ITISE 2018

International Conference on Time Series and Forecasting

PROCEEDINGS OF PAPERS

Volumen 2

ITISE 2018
International Conference on Time Series and Forecasting

Proceedings of Papers
19-21 September 2018
Granada (Spain)

Editors and Chairs
Olga Valenzuela
Fernando Rojas
Héctor Pomares
Ignacio Rojas

I.S.B.N: 978-84-17293-57-4
Legal Deposit: Gr 1165-2018
Edit and Print: Godel Impresiones Digitales S.L.

Preface

We are proud to present the set of final accepted papers for the fourth edition of the ITISE 2018 conference "International work-conference on Time Series" held in Granada (Spain) during September, 19-21, 2018.

The ITISE 2018 (International work-conference on Time Series) seeks to provide a discussion forum for scientists, engineers, educators and students about the latest ideas and realizations in the foundations, theory, models and applications for interdisciplinary and multidisciplinary research encompassing disciplines of computer science, mathematics, statistics, forecaster, econometric, etc, in the field of time series analysis and forecasting.

The aims of ITISE 2018 is to create a friendly environment that could lead to the establishment or strengthening of scientific collaborations and exchanges among attendees, and therefore, ITISE 2018 solicits high-quality original research papers (including significant work-in-progress) on any aspect time series analysis and forecasting, in order to motivating the generation, and use of knowledge and new computational techniques and methods on forecasting in a wide range of fields.

The list of topics in the successive Call for Papers has also evolved, resulting in the following list for the present edition:

1. **Time Series Analysis and Forecasting**.
    - Nonparametric and functional methods
    - Vector processes
    - Probabilistic Approach to Modeling Macroeconomic Uncertainties
    - Uncertainties in forecasting processes
    - Nonstationarity
    - Forecasting with Many Models. Model integration
    - Forecasting theory and adjustment
    - Ensemble forecasting
    - Forecasting performance evaluation
    - Interval forecasting
    - Econometric models
    - Econometric Forecasting
    - Data preprocessing methods: Data decomposition, Seasonal adjustment, Singular spectrum analysis, Detrending methods, etc.
2. **Advanced method and on-Line Learning in time series**.
    - Adaptivity for stochastic models
    - On-line machine learning for forecasting
    - Aggregation of predictors
    - Hierarchical forecasting
    - Forecasting with Computational Intelligence
    - Time series analysis with computational intelligence

- Integration of system dynamics and forecasting models

3. **High Dimension and Complex/Big Data**.

   - Local Vs Global forecast
   - Techniques for dimension reduction
   - Multiscaling
   - Forecasting Complex/Big data

4. **Forecasting in real problem**.

   - Health forecasting
   - Telecommunication forecasting
   - Modelling and forecasting in power markets
   - Energy forecasting
   - Financial forecasting and risk analysis
   - Forecasting electricity load and prices
   - Forecasting and planning systems
   - Real time macroeconomic monitoring and forecasting
   - Applications in: energy, finance, transportation, networks, meteorology, health, research and environment, etc.

After a careful peer review and evaluation process (each submission was reviewed by at least 2, and on the average 3.2, program committee members or additional reviewer). In this proceedings we are presetting the abstract of the contribution to be presented during ITISE-2018 (accepted for oral, poster or virtual presentation,according to the recommendations of reviewers and the authors' preferences).

In this edition of ITISE, we are honored to have the following invited speaker:

1. Prof. Dr. Peter M Robinson , Tooke Professor of Economic Science and Statistics Department of Economics, London School of Economics .

2. Prof Andrew C. Harvey, Emeritus Professor of Econometrics in the Faculty of Economics, University of Cambridge, and a Fellow of Corpus Christi College.

3. Prof. Salah Bourennane, Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, Marseille, France.

4. Dr Karsten Webel, Deutsche Bundesbank, Central Office, Directorate General Statistics Germany.

5. Prof. Dr. Robert Kunst, Professor of Economics at the University of Vienna and affiliated with the IHS (Institute for Advanced Studies) .

6. Prof. Dr. Uwe Hassler, Applied Econometrics and International Economic Policy. Goethe University Frankfurt .

During ITISE 2018 several Special Sessions will be carried out. Special Sessions will be a very useful tool in order to complement the regular program with new and emerging topics of particular interest for the participating community. From the organization of ITISE, we would like to thank deeply the great work that the organizers of Special Sessions do. Thank you very much for your great effort and interest.

Special Sessions that emphasize on multi-disciplinary and transversal aspects, as well as cutting-edge topics are especially encouraged and welcome. and in this edition of ITISE 2018 are the following:

1. *Forecasting Evolution*, Prof. Philip Gerrish, School of Biology, Georgia Institute of Technology, 310 Ferst Dr, Atlanta, GA 30332 .

2. *Forecasting Climate Weather and Operation Impact on Reliability, Safety and Resilience of Critical Infrastructures*, Prof. Krzysztof Kolowrocki, Gdynia Maritime University, Poland, and Prof. Joanna Soszynska-Budny, Gdynia Maritime University, Poland

3. *Applications of time series for hydro-climatic data*, Prof. Bruno Remillardi, Professor at HEC Montral. Consultant at the National Bank of Canada and Prof. Bouchra R. Nasri .

4. *Times series analysis in geosciences*, Prof. Eulogio Pardo-Igzquiza, Professor at Instituto Geolgico y Minero de Espaa (IGME) and Prof. Francisco Javier Rodrguez-Tovar, Depart. Estratigrafa y Paleontologa, University of Granada, Spain.

5. *Forecasting in High Dimension and Complex/Big Data* , Prof. Dr. Luis Javier Herrera and Prof. Dr. Ignacio Rojas , Dep. Computer Architecture and Computer Technology, University of Granada, Spain

6. *Quantum Computing*, Prof. Peter Gloesekoetter, Fachbereich Elektrotechnik und Informatik, Stegerwaldstrae 39, 48565 Steinfurt, Germany. and Dr. Bernd Burchard, Elmos Semiconductor AG, Germany.

7. *Computational Intelligence methods for Time Series*, Prof. Dr. Hctor Pomares , Dep. Computer Architecture and Computer Technology, University of Granada, Spain and Prof. Dr. German Gutierrez , Dep. Computer Science, E.P.S. University Carlos III of Madrid, Spain

8. *Structural Time Series Models*, Prof. Dr. Fernando Rojas , Dep. Computer Architecture and Computer Technology, University of Granada, Spain

9. *Recent Developments on Time-Series Modelling*, Prof. Dr. Olga Valenzuela, Applied Mathematics, University of Granada, Spain

10. *Expert Systems with Time Series - Data*, Prof. Dr. Kalle Saastamoinen , Department of Military Technology, National Defence University,Helsinki, Finland

11. *Spatio-temporal brain dynamics in attention tasks*, Prof. Dr. Juan Manuel Grriz , University of Granada, Spain, and Prof. Dr. Pedro A. Valdes-Sosa , Cuban Neurosciences Center and Prof. Dr. Csar Germn Castellanos Dominguez , Universidad Nacional de Colombia

This new edition of ITISE was organized at the Universidad de Granada, with the help of the Spanish Chapter of the IEEE Computational Intelligence Society and Spanish Network Time

Series (RESET). We wish to thank to our main sponsor the institutions Faculty of Science, Dept. Computer Architecture & Computer Technology and CITIC-UGR from the University of Granada for their support. We wish also to thank to the Dr. Veronika Rosteck and Dr. Eva Hiripi, Springer, Associate Editor, for their interest in the future editing a book series of Springer from the best papers of ITISE 2018.

We would also like to express our gratitude to the members of the different committees and to the reviewer for their support, collaboration and good work.

September, 2018
Granada

ITISE Editors and Chairs
Olga Valenzuela
Fernando Rojas
Hector Pomares
Ignacio Rojas

Program Committee

Bahram Abediniangerabi	University of Texas at Arlington
Sajjad Ahmad	University of Nevada, Las Vegas
Dorel Aiordachioaie	University Dunarea de Jos of Galati
Jose M. Amigo	Universidad Miguel Hernandez
Josu Arteche	University of the Basque Country UPV/EHU
Marcel Ausloos	GRAPES
José Luis Aznarte M.	Artificial Intelligence Department - UNED
Rosangela Ballini	IE - DTE - UNICAMP
Yukun Bao	School of Management, Huazhong University of Sci.&Tech.,
Ildar Batyrshin	Instituto Politecnico Nacional
Salah Bourennane	Institut Fresnel
Bernd Burchard	ELMOS
German Castellanos	Universidad Nacional de Colombia
João P. S. Catalão	University of Porto
Lee Chang-Yong	Kongju National University
Fuxia Cheng	Illinois State University
Paulo Cortez	University of Minho
Pierpaolo D'Urso	Sapienza University of Rome
Ricardo de A. Araújo	Laboratório de Inteligência Computacional do Araripe / Instituto Federal do Sertão Pernambucano
Francisco Estrada	1) Centro de Ciencias de la Atmósfera Universidad Nacional Autónoma de México; 2) Institute for Environmental Studies, VU University Amsterdam
Peter Gloesekoetter	Muenster University of Applied Sciences
Jesus Gonzalo	U. Carlos III de Madrid
Alberto Guillen	University of Granada
German Gutierrez	University Carlos III
Juan M. Gálvez	University of Granada
Ferda Halicioglu	Istanbul Medeniyet University
Marc Hallin	Université libre de Bruxelles
Uwe Hassler	Goethe University Frankfurt
Luis Herrera	University of Granada
Tzung-Pei Hong	Department of Computer Science and Information Engineering, National Univesity of Kaohsiung
Wei-Chiang Hong	School of Education Intelligent Technology, Jiangsu Normal University, China
Samrad Jafarian-Namin	Yazd University
Vinayakam Jothiprakash	none
Sreekanth K J	KISR
Rebecca Killick	Lancaster University
Alexey Koronovskiy	Saratov State University, Faculty of Nonlinear Processes
Dalia Kriksciuniene	Vilnius University
Clifford Lam	London School of Economics and Political Science
Elmar Lang	University of Regensburg
Hooi Hooi Lean	Universiti Sains Malaysia
Junsoo Lee	University of Alabama

Chunshien Li	Department of Information Management, National Central University
Carlos Lima	University of Brasilia
Hui Liu	Central South University, China & University of Rostock, Germany
Wieslaw M. Macek	Space Research Centre, Polish Academy of Sciences
Francisco Martínez-Álvarez	Universidad Pablo de Olavide
Anke Meyer-Baese	FSU
Janusz Miśkiewicz	University of Wrocław
Antonio Montañés	University Zaragoza
Miquel Montero	Universitat de Barcelona
Nageswara Rao Moparthi	V R SIDDHARTHA ENGINEERING COLLEGE
Fionn Murtagh	University of Huddersfield
Guy Mélard	Université libre de Bruxelles
P. C. Nayak	National Institute of Hydrology
Juan M. Palomo-Romero	University of Córdoba
Eulogio Pardo-Iguzquiza	Geological Survey of Spain (IGME)
Eros Pasero	Politecnico di Torino
Fernando Perez De Gracia	Universidad de Navarra
Irina Perfilieva	University of Ostrava
Hector Pomares	University of Granada
María Dolores Pérez Godoy	Departamento de Informática. Universidad de Jaén
Vadlamani Ravi	IDRBT, Hyderabad
Bruno Remillardi	HEC Montreal
Antonio Jesús Rivera Rivas	Departamento de Informática. Universidad de Jaén
Paulo Rodrigues	Banco de Portugal
Ignacio Rojas	University of Granada
Heather Ruskin	Dublin City University
Gerhard Rünstler	European Central Bank
Kalle Saastamoinen	National Defence University of Finland
Leonid Sheremetov	Mexican Petroleum Institute
Ansgar Steland	RWTH Aachen University
Yixiao Sun	University of California San Diego
Leopold Sögner	Institute for Advanced Studies
Ryszard Tadeusiewicz	AGH University of Science and Technology, Krakow, Poland
Chor Foon Tang	Universiti Sains Malaysia
Alicia Troncoso	Universidad Pablo de Olavide
Mehdi Vafakhah	Tarbiat Modares University
Olga Valenzuela	University of Granada
Dimitris Varoutas	National and Kapodistrian University of Athens, Facutly of Informatics & Telecommunications
Claudia Villalonga	Universidad Internacional de La Rioja
Michael Wolf	University of Zurich
Ruqiang Yan	Southeast University
Gilney Zebende	UEFS
Wei-Xing Zhou	ECUST

Table of Contents

Energy Forecasting

Real macroeconomic monitoring and forecasting (Part. I)

Atmospheric Science Forecasting

Advanced econometric methods

Health Forecasting

Econometric models (Part.I)

Computational Intelligence methods for Time Series

Nonstationarity Time Series

Real Macroeconomic Monitoring and Forecasting (Part.II)

Advanced methods in Forecasting

Econometric models (Part.II)

Advanced in Time Series and Forecasting. Virtual Presentation

Outlier Identification in Multivariate Time Series: Boilers Case Study

Joana Ribeiro[1,2][0000−0003−3049−5473], Mário Antunes[2][0000−0002−6504−9441], Diogo Gomes[2][0000−0002−5848−2802], and Rui L. Aguiar[2][0000−0003−0107−6253]

[1] DETI, Universidade de Aveiro - Aveiro, Portugal
joanapribeiro@ua.pt
[2] Instituto de Telecomunicações, Universidade de Aveiro - Aveiro, Portugal
{joanapribeiro,mario.antunes,dgomes,ruilaa}@av.it.pt

Abstract. The existence of abnormal values in data sets of day-to-day actions is common. Denoted as outliers, these values are frequently excluded from the dataset. However, in fraud detection, disease diagnosis and other areas, outliers often add relevant knowledge. In this paper, we present an application of a methodology to detect outliers in multivariate time series by translating the multi-type variables into textual data (strings). After that, classification algorithms are applied. An application is performed into a real dataset that represents boiler operations and the main goal is to identify their faults. A major part of the work concerns the dataset processing steps that enable the application of machine learning algorithms. Besides the boiler malfunctions, normal operation cycles were also identified. We aim to improve the real-time fault identification of the operating devices allowing safer appliances.

Keywords: Outlier Prediction · Multivariate Time Series · Machine Learning

1 Introduction

With the advances in technology, an increasing number of datasets has been generated from real-life events and made available for further study by industry and academics. Machine Learning has gain traction as a novel method to process these datasets and predict relevant patterns and knowledge. For example semantic [2] and stream features [10] are being used to tackle the massive amount of data generated by IoT/M2M scenarios. However, machine learning methods do not deal well with sequential data - time series. Instead of unchangeable variables, such as petal length and width in the Iris flower dataset[3], timeseries datasets are a complex set of variables that evolve through time.

Abnormal values are common in real-life datasets. Frequently denoted as outliers, they are different from what is considered to be normal in each specific case study. In some situations, these values are not common dataset noise, but

[3] https://archive.ics.uci.edu/ml/datasets/iris

instead they represent changes or malfunctions in the systems [3,8,9,17,18,21]. Considering environmental changes or fraud detection, outliers provide useful knowledge. Machine learning methods are used to detect abnormal values, each in turn can be used to improve human intervention, for example when predicting components faults.

The dataset in this paper contains boilers operational data. The goal is to identify the abnormal operation/state through the classification of the cycles. Since we are considering real data, normal operation is more frequent than failure. This leads to an imbalanced dataset due to the low frequency of outliers, which usually translates into a classification preference for the most frequent class since most machine learning algorithms are statistically based.

Besides the time variable, the dataset is rather complex due to the variety of variables from numeric to textual (strings). In order to deal with this real-world dataset we used the value-trend approach: a way of representing time series by textual data. This transformation allows us to apply some common machine learning algorithms and process the time series as a classification problem.

The remainder of the paper is organized as follows. In Section 2, the main characteristics of the dataset are presented along with the first pre-processing steps. Section 3 concerns the value-trend approach along with its application to the dataset. In Section 4, as an example of application, one classification algorithm is constructed considering some set variations and the performance measures are investigated. Conclusions and future work are presented in Section 5.

2 Dataset

As stated before, the dataset in this paper contains boilers operation values. These boilers, whose main function is to supply hot water, where installed in customer's houses. The hot water can be used as domestic water (hot water – HW – cycle), such as hot water for bathing or kitchen use, or as central heating (central heating – CH – cycle), in which hot water circulates along the house into wall-hung heating devices. There is an additional mode called boost cycle, which provides a quicker supply of hot water. These cycles are considered normal, and are the most representative data patterns.

Boilers have sensors collecting data such as: temperatures, water flow, number of heating requests, state variables such as open or closed valves, among others. Therefore, the dataset is constituted by continuous and discrete variables. State variables can be transformed in discrete values since they always assume two opposite values: On/Off, Yes/No, HW/CH, among others, and so, these variables were binary coded. The dataset is constituted by 40 discrete and 29 continuous variables, along with four textual values concerning the boiler model and the gateway software.

The last variable is the fault code. Malfunctions are associated with faults identified by the appliance software and automatically tagged with a fault code. Currently there are 39 different fault codes. The dataset is composed by labelled

Table 1. Anonymize data.

ID	Date	3w Valve	Power	CH Flame	Flow	Power	Fault
123XYZ	y-m-d h:m:s.ms	Hot_water	0	No Off	0	0	/200
123XYZ	y-m-d h:m:s.ms	Hot_water	44	Yes Off	9.5	100	D4/271
123XYZ	y-m-d h:m:s.ms	Cent._heat.	78	Yes On	12.4	153	E2/21

and unlabelled data: known malfunctions (classified by the software with a fault code), but also malfunctions that were unclassified (unlabelled data). Our goal is to use classification algorithms to identify the fault codes of the unlabelled data. Novelty detection [15] is one possible method for outlier identification. It models the correct behaviour of a system, and labels all the behaviour that do not fit in such patterns as outliers.

Unfortunately normal operation cycles are not identified by the appliance software and consequently, they are not labelled in the dataset. To train classification algorithms, we must attribute a class to all the observations of (at least) the train dataset. A few variables let us recognize if there has been any hot water or central heating request. Therefore, 100 observations of each HW, CH and boost cycles were manually labelled considering expected behaviours. There are no consensual rules to decide if a cycle is of normal operation or not.

Through data visualization of normal cycles and classified faults, it is possible to identify variables with no classification power, such as those that do not vary between labels. This allow us to reduce the dataset to 40 variables. Since not all boilers have the same sensors, we were able to reduce to a total of 31 variables instead of the original 74.

The data was collected within one year and four months concerning 1563 appliances of about five different models and stored in MongoDB[4] database. Although the data is updated at every millisecond, only the variables for which there has been any change in the value or state are recorded. As a consequence, the dataset is riddle with missing entries, which do not always correspond to missing vales but rather omitted values. Consider the reduced example of Table 2 concerning three variables during 60 milliseconds of operation. Investigating the variable "Water Temp.", the value 20 has been recorded at 01:30:00.000. Since there were no value changes for this variable in the next milliseconds, the following entries are empty (lines two, three and four of the second column) although the appliance is always recording. At 01:30:00.400 a new value has been detected: 22, and so, a new entry was recorded for the "Water Temp." variable.

One way to deal with omitted values would be to copy the last saved value until a new entry happens, for each column of the dataset, one by one. However, this leads to a fake translation of the boiler operation. There is one specific fault – data loss due to connectivity problems of the gateway – that could be hidden by this method. When the gateway looses connectivity, data is not transmitted. So, copying the data would not allow to identify this fault, since there would be

[4] `https://www.mongodb.com/`

Table 2. Data recording example.

Time	Water Temp.	Gas	Fan
01:30:00.000	20		
01:30:00.100		1	
01:30:00.200			1
01:30:00.300			0
01:30:00.400	22		
01:30:00.500	23	0	

no missing values. Also, a subset of distinct faults is due to different sensors that stop working. For example, if some temperature sensor breaks, the temperature is not measured until the sensor is fixed. By copying the last saved value until a new entry happens, it would give the idea of a stable temperature over time. However, the temperature could have changed due to some heat request. Using this data copying technique, the sensor failure would be impossible to identify, since no missing data would exist. Given our dataset, through data analysis it is possible to conclude that state variables always register a new entry around every four minutes, even if there has been no state change. Therefore, it is assured that there was no connectivity faults through these four minutes. So, the continuity of values was accomplished by verifying if, for any predictive variable, the last saved value was within less than four minutes. If that happens, then the last value of each column is copied until a new value appears. This allows the identification of the data loss fault, since it happens for more than four minutes, and so, in this case there are no registers of the state variables for more than four minutes. However, in the case of sensor faults it is impossible to avoid such errors. There is an exception that happens when, for example in the temperature sensor fault, the variable concerning that temperature doesn't have saved values in the matrix time range. Therefore, there would be no exist values to copy through that column. So, for such few cases, sensor faults can be identified by data loss (empty columns in the matrix), together with other data patterns.

3 Value-trend Approach

The main difficulty of this problem is in the several specificities of the multivariate time series that machine learning algorithms have difficulty to deal with. Some variables are represented as strings, like the boiler name, while others as numerical values, such as temperatures. Also, it is very important to capture each variable behaviour and relate all the variables between them and time. Several algorithms have been proposed to deal with time series data [4,6,7,11,12,16,19], but we could not find any able to solve our specific problem due to the high specificity of the data.

The algorithm studied is responsible for representing the time series in a simpler and unified way, enabling the implementation of machine learning al-

$$\mathbf{TS} = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1k} \\ x_{21} & x_{22} & ... & x_{2k} \\ ... & ... & ... & ... \\ x_{m1} & x_{m2} & ... & x_{mk} \end{bmatrix} \Rightarrow \mathbf{PAA} = \begin{bmatrix} < w_{11}, w_{12}, ..., w_{1n} > \\ < w_{21}, w_{22}, ..., w_{2n} > \\ ... \\ < w_{k1}, w_{k2}, ..., w_{kn} > \end{bmatrix}^{\top}$$

Fig. 1. Application of the PAA algorithm in matrix format.

gorithms. It was devised by Eamonn Keogh and Jessica Lin in 2002 [14] and consists of two main steps: the transformation of the time series into sets of vectors through the piecewise aggregate approximation algorithm (PAA) and the conversion of those vectors into a set of letters by the symbolic aggregate approximation algorithm (SAX). An additional phase where a trend analysis is performed as an improvement of the SAX algorithm is also discussed. This data transformation is made considering each predictive variable individually. Our novel approach consists in apply existent methodologies into a multivariate, wide-variable (in terms of the types of variables) time series, so machine learning techniques can than be used to classify the time series.

3.1 Piecewise Aggregate Approximation (PAA)

The PAA algorithm divides the time series into a vector of equally sized segments [1,20]. By definition, this algorithm transforms any time series X of length m into n segments of time, resulting in a vector $X = \{x_1, x_2, ..., x_n\}$ of temporal segments, where n is any arbitrary integer such that $n \leq m$. Then, considering each of the segments, the average value of each variable is calculated: for each temporal segment i, with $i = \{1, 2, ..., n\}$, and each predictive variable j, the mean value w_{ji} is obtained.

The number of segments can vary between one segment, and the entire time series is equal to as many univariate vectors as the number of variables, up to the number of value registrations considered for the time series, taking each vector the same dimension as the time series. In the case of our data set, after applying the PAA algorithm, we will have the same number of vectors as the number of predictive variables. Each vector will have dimension equal to the chosen number of segments. Thus, each time series is represented by a set of multidimensional vectors, all with the same dimension.

Consider the time series present in the left of Figure 1, represented as a matrix of m lines and k columns, corresponding to m time registrations of k variables. After applying the PAA algorithm, we obtain the matrix of the right side of the same Figure. Each w_{ji} represents the mean value of the $j-th$ variable in the $i-th$ segment, with $j = \{1, 2, ..., k\}$ and $i = \{1, 2, ..., n\}$. Note that we are considering the partition of the time series into n segments and, so, we have a dimension reduction.

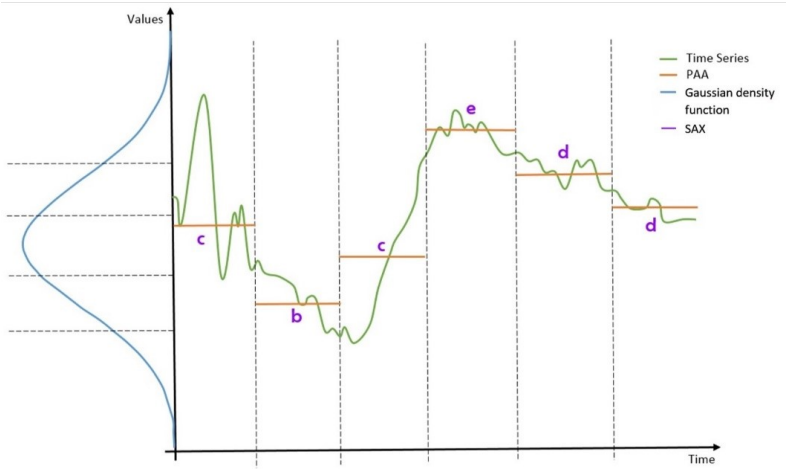


Fig. 2. Application of SAX algorithm in a univariate time series.

3.2 Symbolic Aggregate Approximation (SAX)

The SAX algorithm consists in assigning a letter to each mean value previously obtained with the PAA algorithm [14].

An extensive and rigorous analysis performed in [13] has shown that time series data, after being normalized by Z-score, usually follows a Gaussian distribution. This detail enables a conscious attribution of letters to the mean values, and allows the partition of the Gaussian probability density function into equally spaced break-points. This partition must be performed according to the number of letters that we want to associate with the mean values. By assigning a letter to each area obtained by partitioning the probability density function, each mean value gets its own letter. Figure 2 shows a time series with one normalized variable, where six segments were considered for the PAA algorithm. Then, four breakpoints were defined in the Gaussian probability density function, resulting in the partition of possible mean values into five intervals. Subsequently, a letter of the alphabet was associated with each mean value. For example, the mean value in the second segment is above the first break-point and below the second one defined by the Gaussian function. So, this segment has been associated with the letter 'b'. Proceeding successively in this way, the time series represented as a vector of mean values, from the PAA algorithm, is now transformed into a string. In this case, the resulting string representative of the time series is 'cbcedd'.

3.3 Trend Analysis

A study carried out in [5] suggests that after applying the SAX algorithm a new and additional phase should be performed. The idea is to associate each temporal segment not only with a letter (representing the mean value) but also

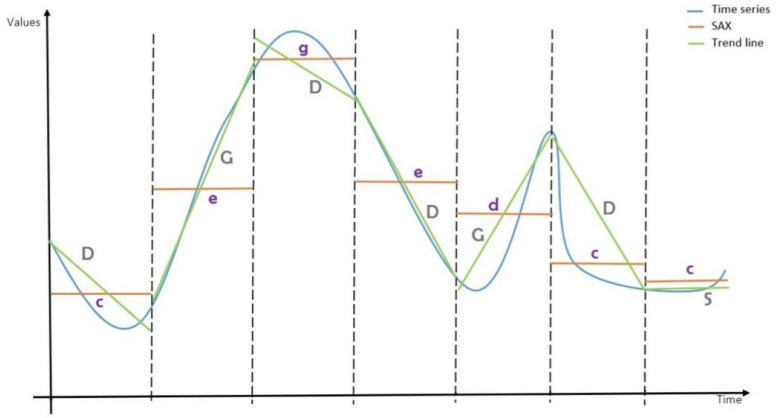


Fig. 3. Value-Trend approach applied to a univariate time series.

its tendency. In fact, we can easily obtain two different time series with the same SAX representation but, when associating its trends, becoming distinct strings. So, the objective is to capture the trend in each temporal segment constructed in the PAA algorithm.

The attribution of the trend information is made through its association to a straight line. The search for the line that best fits each variable behaviour is performed using the least squares method, that is, we intend to obtain the model $y = ax + b$ such that $\sum_{t-1}^{s} = (y_t - (ax_t + b))^2$ is minimized, where x denotes the time variation, with $t = 1, ..., s$ considering the s existing values registrations of the variable in the segment under transformation, and y represents the value of the variable in each time point t. Note that this step is performed for each segment, similarly to the steps taken in the SAX algorithm. Then, after the line is obtained, its slope is used to assign a second letter to the variable under study in this time segment: 'G' for growth, 'D' for decay or 'S' for stable.

Consider the univariate time series represented in Figure 3 where after the application of the SAX algorithm with seven temporal segments and seven letters to represent the mean values, we obtain the sequence 'cegedcc', which is represented in purple colour. Then, a straight line, represented by the green colour, was adjusted to the variable in each temporal segment. According to the line slope, a second letter was attributed to each temporal segment which is represented with a grey colour. In the end, the complete time series is represented by the string 'cDeGgDeDdGcDcS'.

4 Implementation and Evaluation

After processing the data, we apply the value-trend approach followed by the classification methods. All the previous and next steps were performed using Matlab software. First, the value-trend approach was implemented. Then, as an

example of application, some decision tree models were constructed and evaluated taking some variations of the data set into consideration.

Main SAX code for Matlab is available for download[5]. The output of the SAX algorithm for each predictive variable is a string representing the mean value of the variable in each segment. The length of this string is equal to the number of considered segments. Some modifications were made to the original code, so we can also obtain the trend approximation. The segments are the same that were considered in the SAX algorithm. "polyfit" Matlab function was used to obtain the slope of the wanted straight line. The modified algorithm output is the string from the value-trend representation of each predictive variable.

The choice of the number of segments was based on each time series length. For data relative to more than ten minutes, the time series is divided into segments of two minutes. For time series with length between two and ten minutes, segments of 30 seconds have been formed. In the case of time series with less than two minutes, the number of segments is equal to twenty times the time series length in minutes.

The appliance's software can identify 39 different faults, each fault related with a specific component failure and possible maintenance solutions documented. Thus, a total of 42 classes were considered for the time series classification problem, corresponding to the three normal operation cycles and the 39 existing fault codes.

When trying to understand the appliances faults, we found that the causes approximate to some simple logic rules. For example, consider that there is an order to use hot water but the temperature at the output is not increasing, despite it is below the set point. This fault can be translated into: hot water order $\Rightarrow$ temperature not increasing and temperature below the set point $\Rightarrow$ fault. So, the decision tree model was the first one making sense to be used for the classification problem as it is also based on simple logic rules. The decision model construction was made considering 10 fold cross validation and the Matlab command "fitctree". Three different splitting algorithms – "pull left by purity", "principal component-based partitioning" and "one versus all by class" – were used to construct the decision rules, all of them available in Matlab. Also, Matlab skills were used to optimize some parameters of each model, such as the minimum number of observations in each leaf node or the maximum number of splits. The choice of the best parameters values is based on the minimization of the cross validation loss.

After testing the splitting rules, "one versus all by class" was the one that minimized the cross validation loss. A new decision tree model was then constructed considering this splitting rule and its optimized parameters. In this next phase, we used 5 folds cross validation instead of 10 folds due to the low frequency of occurrences per class. Measures of performance are presented in Table 3, remarking that the obtained accuracy is considerably low (38.46%).

The number of labelled occurrences per fault varies between 1 to 133 in a total of 1185 classified cycles. As is possible to see in Figure 4, this is clearly

[5] https://cs.gmu.edu/ jessica/sax.htm

Table 3. Performance measures for the optimized decision tree model in the 42 class problem.

Measures	
Classified Observations	99.83%
Accuracy	38.46%

an unbalanced data set. We note that the low accuracy result is related to the multi-class problem with 42 classes. Therefore, a new approach was made: the labels were reduced to the classes that present more occurrences than the mean value of occurrences – 11 classes – plus the normal cycles – 3 classes each with 100 occurrences – in a total of 14 classes.

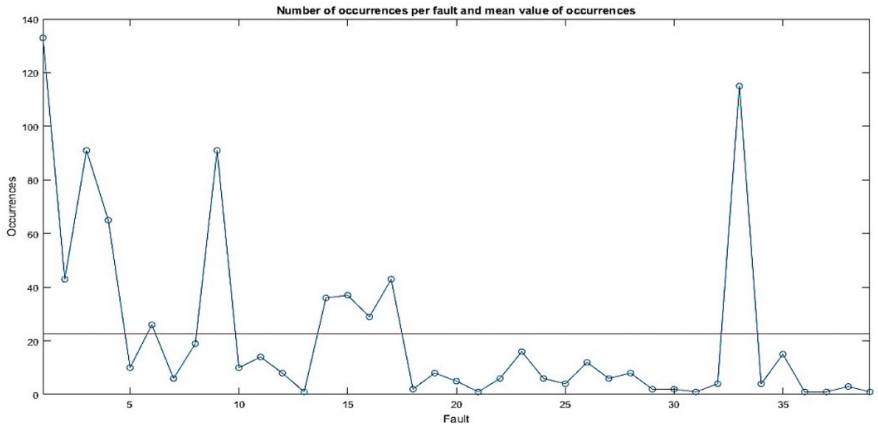


Fig. 4. Number of occurrences per fault and mean value of occurrences (horizontal line).

The decision tree models were constructed once again considering 10 folds cross validation, the previously tested splitting rules and using the optimization of parameters option. Once again, the "one versus all by class" splitting rule is the one that minimizes the cross validation loss. The accuracy measures were obtained with a 5 fold cross validation over the ptimized model (see Table 4). Although the accuracy remains low, the results were better than considering the 42 class problem. Also, in the 14 class problem, the observations were all classified, when before only 99.83% were attributed to one class.

Finally, the classes were reduced to two: fault or normal cycle. The idea is to evaluate the appliances with more confidence, since it is not always needed to know which fault occurred but just the frequency of faults over time (allowing to study the appliance lifetime). A new decision tree model was constructed and optimized, using the previously mentioned options. In this case, the "pull left by

Table 4. Performance measures for the optimized decision tree model in the 14 class problem.

Measures	
Classified Observations	100.00%
Accuracy	43.30%

purity" splitting rule was the one that minimize the cross validation loss. The new model was evaluated using 5 and 10 fold cross validation. As expected, the results were far better compared with the previous models (see Table 5).

Table 5. Performance measures for the optimized decision tree model in the 2 class problem.

Measures	Classified Obs.	Sensitivity	Specificity	Accuracy	Precision	F1
5 folds	100.00%	79.00%	95.48%	91.31%	85.56%	82.15%
10 folds	100.00%	60.33%	99.55%	89.62%	98.22%	74.75%

The specificity was in both cases higher than the sensitivity. Since in this study the "positive" class corresponds to the normal cycles and the "negative" class to the faults, the algorithm is better at classifying faults than normal cycles. The decision tree model trained with 5 folds has classified the true class more times than the 10 folds decision tree models, according to the accuracy results, while the precision of the results were better in the 10 folds trained model – and so, this is a more consistent model. F1-measure reflects the combination of these two last metrics. Consequently, better results were obtained when considering the 5 folds cross validation technique, which is expected since this problem suffers from number of observations.

5 Conclusions

The main difficulty of this work was to find an algorithm able of classify time series with predictive variables in the form of strings, continuous and discrete values. The final solution was to transform the time series using the value-trend approach, where each predictive variable was coded as one string. In this way, it was possible to express all the different types of predictive variables into just one (string). Also, the behaviours previously represented in a set of matrix lines were transformed into a single string (one line matrix). Therefore, each time series became a single matrix of strings, and so, a multivariate classification problem of string type variables.

Since the behaviour is the most important feature of the problem at hand, the trend variation was also considered to extract information from the time series. This is a novel approach considered in [5] to the existent SAX algorithm.

The main goal of this work was the construction of an algorithm able to identify faults, and so, the 5 folds decision tree model is the more useful model. Also, not only the faults were identified but also cycles of normal operation.

Future work includes the prediction of faults one week in advance. Also, it is important to distinguish the faults of the sensors that were not possible to obtain due to the continuity of values approach. The novelty detection family of methods should also be put into comparison with the presented approach.

Acknowledgement

The present study was developed in the scope of the **Smart Green Homes** Project [POCI-01-0247-FEDER-007678], a co-promotion between **Bosch Termotecnologia S.A.** and the **University of Aveiro**. It is financed by Portugal 2020 under the Competitiveness and Internationalization Operational Program, and by the European Regional Development Fund.

References

1. Alsallakh, B., Bögl, M., Gschwandtner, T., Miksch, S., Esmael, B., Arnaout, A., Thonhauser, G., Zöllner, P.: A visual analytics approach to segmenting and labeling multivariate time series data. In: EuroVis Workshop on Visual Analytics (EuroVA). The Eurographics Association (2014). https://doi.org/10.2312/eurova.20141142
2. Antunes, M., Gomes, D., Aguiar, R.L.: Towards IoT data classification through semantic features. Future Generation Computer Systems (dec 2017). https://doi.org/10.1016/j.future.2017.11.045
3. Chatzigiannakis, V., Papavassiliou, S., Grammatikou, M., Maglaris, B.: Hierarchical anomaly detection in distributed large-scale sensor networks. In: 11th IEEE Symposium on Computers and Communications (ISCC06). IEEE (2006). https://doi.org/10.1109/iscc.2006.1691116
4. Cui, Z., Chen, W., Chen, Y.: Multi-scale convolutional neural networks for time series classification. CoRR **abs/1603.06995** (2016), `http://arxiv.org/abs/1603.06995`
5. Esmael, B., Arnaout, A., Fruhwirth, R.K., Thonhauser, G.: Multivariate time series classification by combining trend-based and value-based approximations. In: Computational Science and Its Applications – ICCSA 2012, pp. 392–403. Springer Berlin Heidelberg (2012). https://doi.org/10.1007/978-3-642-31128-4_29
6. Formisano, E., Martino, F.D., Valente, G.: Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. Magnetic Resonance Imaging **26**(7), 921–934 (sep 2008). https://doi.org/10.1016/j.mri.2008.01.052
7. Górecki, T., Łuczak, M.: Multivariate time series classification with parametric derivative dynamic time warping. Expert Systems with Applications **42**(5), 2305–2312 (apr 2015). https://doi.org/10.1016/j.eswa.2014.11.007
8. Hassan, A., Mokhtar, H.M., Hegazy, O.: A heuristic approach for sensor network outlier detection. International Journal of Research and Reviews in Wireless Sensor Networks (IJRRWSN) (2011)

9. Janakiram, D., Reddy, V., Kumar, A.: Outlier detection in wireless sensor networks using bayesian belief networks. In: 2006 1st International Conference on Communication Systems Software & Middleware. IEEE (2006). https://doi.org/10.1109/comswa.2006.1665221
10. Jesus, R., Antunes, M., Gomes, D., Aguiar, R.L.: Modelling patterns in continuous streams of data. Open Journal of Big Data (OJBD) **4**(1), 1–13 (2018), `http://nbn-resolving.de/urn:nbn:de:101:1-201801234777`
11. Kadous, M.W.: Learning comprehensible descriptions of multivariate time series. In: In Ivan Bratko and Saso Dzeroski, editors, Proceedings of the 16 th International Conference of Machine Learning (ICML-99. pp. 454–463. Morgan Kaufmann (1999)
12. Kadous, M.W., Sammut, C.: Classification of multivariate time series and structured data using constructive induction. Machine Learning **58**(2-3), 179–216 (feb 2005). https://doi.org/10.1007/s10994-005-5826-5
13. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery - DMKD03. ACM Press (2003). https://doi.org/10.1145/882082.882086
14. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series. Data Mining and Knowledge Discovery **15**(2), 107–144 (apr 2007). https://doi.org/10.1007/s10618-007-0064-z
15. Pimentel, M.A., Clifton, D.A., Clifton, L., Tarassenko, L.: A review of novelty detection. Signal Processing **99**, 215–249 (jun 2014). https://doi.org/10.1016/j.sigpro.2013.12.026
16. Smyl, S., Kuber, K.: Data preprocessing and augmentation for multiple short time series forecasting with recurrent neural networks. In: Conference: 36th International Symposium on Forecasting. Santander (2016)
17. Subramaniam, S., Palpanas, T., Papadopoulos, D., Kalogeraki, V., Gunopulos, D.: Online outlier detection in sensor data using non-parametric models. In: Proceedings of the 32Nd International Conference on Very Large Data Bases. pp. 187–198. VLDB '06, VLDB Endowment (2006), `http://dl.acm.org/citation.cfm?id=1182635.1164145`
18. Van Phuong, T., Hung, L.X., Cho, S.J., Lee, Y.K., Lee, S.: An anomaly detection algorithm for detecting attacks in wireless sensor networks. In: Proceedings of the 4th IEEE International Conference on Intelligence and Security Informatics. pp. 735–736. ISI'06, Springer–Verlag, Berlin, Heidelberg (2006). https://doi.org/10.1007/11760146_111
19. Wang, L., Wang, Z., Liu, S.: An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm. Expert Systems with Applications **43**, 237–249 (jan 2016). https://doi.org/10.1016/j.eswa.2015.08.055
20. Wang, Q., Megalooikonomou, V.: A dimensionality reduction technique for efficient time series similarity analysis. Information Systems **33**(1), 115–132 (mar 2008). https://doi.org/10.1016/j.is.2007.07.002
21. Zhang, Y., Meratnia, N., Havinga, P.: Outlier detection techniques for wireless sensor networks: A survey. IEEE Communications Surveys & Tutorials **12**(2), 159–170 (2010). https://doi.org/10.1109/surv.2010.021510.00088

Realized volatility in the presence of structural breaks: which forecast?

G. Albano, D. De Gaetano

Dipartimento di Scienze Economiche e Statistiche, Università di Salerno, Italy, pialbano@unisa.it
SOSE - Soluzioni per il Sistema Economico S.p.A. Via Mentore Maggini 48/C Roma (00143), ddegaetano@sose.it

Abstract

Forecasting volatility is a main issue for asset pricing, portfolio selection and risk management. In the last three decades a lot of volatility forecast models based on low-frequency data, such as generalized autoregressive conditional heteroskedasticity (GARCH) models and stochastic volatility (SV) models, have been proposed in the literature. As high-frequency data have become available, a new measure of volatility defined as the sum of intraday squared returns, the so-called realized volatility, has been considered. As realized volatility is an observable variable, standard time series models can be used for forecast purposes.

In this paper we focus on some HAR-type models firstly introduced in [3] and reported in [1], consisting essentially in regression models in which the response variable is the realized volatility and different its decompositions are considered as regressors, such as its positive or negative (semivariance) part and its continuous or discontinuous part (jumps).

In practice the considered models can be written in a standard regression framework:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t \tag{1}$$

where y_t is the average realized variance over a fixed period, $\mathbf{x}_t$ is the $p \times 1$ vector of the regressors at time t, relying on the decomposition of realized variance.

However, the financial markets are buffeted by suddenly important events, which can lead to sharp breaks in the markets and thus breaks in parameters governing the volatility models (see, for example, [2]) can occur. In such cases the inference on the parameters may be misleading as well as any policy implications drawn from the model. The accuracy of post-sample forecasting is affected as well.

In particular, assuming that m breakpoints at the date $\tau_1, \tau_2, \cdots, \tau_m$ occur, model (1) can be rewritten as

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_{\tau_{j-1}+1:\tau_j} + \epsilon_t \qquad t = 1, 2, \ldots T \qquad j = 1, 2, \ldots m+1 \tag{2}$$

where, by convention, $\tau_0 = 1$ and $\tau_{m+1} = T$.

The presence of structural breaks can be tested by means of the numerous tests proposed in the econometric literature. Here we use the approach proposed in [4] in which under the alternative H_1 no particular patterns for the changing

parameters are specified. However, once the parameter instability due to the presence of structural breaks has been detected, the problem is how to account for them when generating forecasts.

In practice, the dates of the break points are not known a priori and an estimation procedure has to be used. It could produce imprecise values which affect negatively the specification of the forecasting model and, as a consequence, poor performances of the forecasts. Furthermore, even if the last break date is correctly estimated, the forecasts generated by this scheme are likely to be unbiased and may not minimize the mean square forecast error ([5]). Moreover, if the last break is detected close to the boundaries of the data sample, the parameters of the forecasting model are estimated with relatively short sample and the estimation uncertainty may be large. A possible approach is to consider an estimation window whose size is independent of the last break location. However the forecasting performance is sensitive to the choice of the observation window due to the bias-variance trade off.

In order to solve the problems arising with the choice of a single estimation window, it can be useful to consider forecast combinations generated by the same model but over different estimation windows. A first analysis in this direction has been provided in [6], in which three HAR-type models considering the continuous components of the realized volatility are discussed.

In this paper we extend the results in [6] by considering several alternative and more complex HAR-type models. For each specification we consider some alternative forecast combinations, based on different estimation windows with suitable weighting schemes.

By considering different financial datasets we empirically show that there is a strong evidence of structural breaks in the HAR-type volatility models. Moreover, the considered forecast combinations are able to forecast the realized volatility in the presence of structural breaks, even if their number and the location are unknown. The point forecasting performance of all the considered realized volatility forecast methods are compared in terms of suitable loss functions, by using the Model Confidence Set (MCS) procedure. Since it depends on the choice of the loss function, we choose, in addition to the classical MSE and QLIKE, suitable asymmetric loss functions as suggested in [7] (AMSE). In our analysis , we conclude that forecast combinations taking into account the location of the structural breaks show better performances in terms of AMSE, while the classical MSE and QLIKE seem do not discriminate among the considered forecasters.

References

1. Sevi, B. Forecasting the volatility of crude oil futures using intraday data. Eur. J. Oper. Res. 2014, 235, 643659.
2. Liu, C.; Maheu, J.M. Are there structural breaks in Realized Volatility? J. Financ. Econom. 2008, 6, 326360.
3. Fulvio Corsi 2009. A Simple Approximate Long Memory Model of Realized Volatility. Journal of Financial Econometrics 7(2), 174196,

4. Zeileis, A.; Kleiber, C.; Kramer, W.; Hornik, K. Testing and dating of structural changes in practice. Comput. Stat. Data Anal. 2003, 44, 109123.
5. Pesaran, M.H.; Timmermann, A. Selection of estimation windows in the presence of breaks. J. Econom. 2007, 137, 134161.
6. D. De Gaetano 2018. Forecast Combinations in the Presence of Structural Breaks: Evidence from U.S. Equity Markets. Mathematics 2018, 6(3), 34.
7. G. De Luca, G. Gallo, D. Caritá. Evaluating Combined Forecasts for Realized Volatility Using Asymmetric Loss Functions. Econometric Research in Finance, 2017, 2(2), 99-111.

Experimental Comparison and Tuning of Time Series Prediction for Telecom Analysis

André Pinho (apinho@student.dei.uc.pt), Helena Silva (lena@alticelabs.com), Pedro Furtado (pnf@dei.uc.pt)

University of Coimbra, Altice Labs, Portugal

Abstract. Prediction of consumption is fundamental in telecommunications, for efficient management of network resources, and for guaranteeing quality of service. In this work we investigate the use of time series models to forecast consumption. Two time series forecasting algorithms are compared, Auto-Regressive Integrated Moving Average (ARIMA) and Prophet, launched by Facebook in 2017. We also developed a simple automated parameterization solution for ARIMA, which is important in practical deployments. The work described was developed in the context of tool development effort within Altice Labs that provides actual software to associated Telecom operators, in collaboration with University of Coimbra. To validate results we used real data from a Telecom operator. The forecast results showed that ARIMA was better than Prophet with a Mean Absolute Percentage Error (MAPE) of 3.71% in the three-month forecast and 4.14% in the twelve-month forecast.

Keywords: ARIMA, Forecasting, Prophet

1 Introduction

Currently, telecom operators face competition from other operators and from new services made available through the internet. Operators need to be one step ahead of competition, and they need to offer reliable services to avoid migration of customers and a fall in profits. One important opportunity is to create tools to analyze the huge and valuable data that they collect using data science techniques, with great potential for decision support. In this context, consumption forecasting is critical to provide information that helps the operator efficiently plan and manage network resources and provide an improvement in quality of service.

Consumption forecast in telecommunications presents its challenges, as it is necessary to deal with seasonality, trends, and with the variation of the number of clients. Taking into account these challenges, this work developed at Altice Labs in collaboration with University of Coimbra had two objectives. First, to compare two time series forecasting models, ARIMA and Prophet, in order to verify which one best fits the context of telecommunications. Since ARIMA was superior and since it requires manual parameterization of the model, the second objective was to create an automatic parameterization mechanism. This mechanism consists of a set of steps, selection,

adfa, p. 1, 2011.

transformation, parameterization by exhaustive search, and application of the model chosen in forecasting. With the choice of model and the automated parameterization, the approach is ready to be integrated in decision support tools to be used by managers that do not need to know any details of the forecasting model to use it. Instead, they simply view charts with the forecasts they desire, and the approach adjusts automatically the parameterization to changes in behavior of the data series. In terms of software details, the tool was developed in python, and in particular using its time-series libraries statsmodels [1] and fbprophet [2], containing the ARIMA and Prophet methods respectively.

This paper is divided into 7 sections. Section II presents state of the art. Section III discusses how to apply time series forecasting in the context of telecommunications, reviewing the relevant details of ARIMA and Prophet. In section IV we describe the need for manual parameterization of the time series model, and in section V proposes an automatic parameterization approach for ARIMA using exhaustive search. Section VI reports and analyses our experimental results. Finally, section VII concludes and discusses future work.

2 State of the art

In this section we review works on time-series forecasting in the context of Telecommunications that are most related to ours.

In 2008, S. T., & Sampaio, R. J. B proposed a model to predict short-term consumption of a telecommunications service [3]. Since service consumption presents a non-linear behavior caused by the existence of tendency and seasonality, the authors used two neural network algorithms, the Multilayer Perceptron and the Radial Basis Function network (RBF). The per-month dataset was divided into two sets, training set consisting of a history of 3, 4 and 6 months, and test set consisting of only 1 month ahead. The metric used in the evaluation of the results was Mean Squared Error (MSE). From the experiments performed for different historical periods, the Multilayer Perceptron model presented better prediction quality, although with worse computational performance.

In 2015, Wang, M., Wang, Y., Wang, X., & Wei, Z proposed a model based on the Auto-Regressive Integrated Moving Average (ARIMA), with the objective of predicting performance in telecommunications [4]. In this study they used monthly aggregated data corresponding to periods of two and a half years. The data of the first two years was used for analysis of the time series and for training the model, and the remaining six months were used for validation of the model. The method used to verify the seasonality of the time series was the analysis of the statistical properties, mean, variance and correlation coefficient, verifying if they remain constant over time. In our study it was verified that the series was non-stationary, so it had to be adapted using the typical manual procedure we describe later. The model obtained by [4], had an average error of 1% for five months.

In 2017, Hideaki Hayashi compared the performance of Prophet and ARIMA in the different context of prediction of number of flights in the United States [5].

Prophet was inferior to ARIMA with the parameters configured manually, and superior to ARIMA with the parameters configured automatically with the values by default. The author concluded that the ARIMA method, unlike Prophet, requires manual configuration of the model parameters in order to have good results. This means that the ARIMA requires a lot of knowledge in the domain to be configured manually.

The two works [4] and [5] do show that ARIMA could be the best choice for time-series analysis and forecasting in Telecom and other contexts, but has the big drawback that it requires manual configuration, which is undesirable for integration into a managerial decision support tool as we desired. Furthermore, in our work it was important to evaluate the two alternatives (ARIMA and Prophet) in the context of real Telecom consumption data, and to devise which to integrate into a decision support tool and how to automate its use.

3 Application of time series forecasting in the context of telecommunications

In telecommunications, time series forecasting is typically applied in the forecast of consumption and also in the detection of anomalies in real time. The remainder of this section reviews the concepts of stationarity, the Auto-Regressive Integrated Moving Average (ARIMA) [6] [7] and Prophet [2]. The later was launched by Facebook in 2017 to allow its use by people with less knowledge in the field, since ARIMA requires manual tuning of fundamental parameters.

3.1 Stationarity in ARIMA

An important concept in the application of the ARIMA method is stationarity, since the model can only be constructed with stationary time series. A series is stationary if its statistical properties remain constant over time. The existence of trend and seasonality are two of the reasons that lead the series to be non-stationary [6] [7].

There are two methods that allow you to check whether a series is stationary or not. The first method consists in the graphical visualization of the variation of the statistical properties of the series, such as the moving average (calculation at each instant of the average of the values corresponding to the last seasonal period, typically of twelve consecutive months) and the moving standard deviation over time. If the properties of the series do not change over time, then the series is stationary. The second method, the Dickey-Fuller test, assumes that the null hypothesis is that the series is non-stationary. This test calculates the value of the statistical test and some critical values for different levels of confidence. If the value of the statistical test is less than the critical value, then the series is stationary [6].

Differentiation [6] [7] is one of the existing techniques that allows us to deal with seasonality and trend of the time series, bringing it closer to stationarity in time. At each instant in the series, differentiation subtracts the original observation, Y_t, from that of the previous instant, Y_{t-1}, using the following formula:

$$Y'_t = Y_t - Y_{t-1} \quad (1)$$

3.2 Time series forecasting with ARIMA and with Prophet

This subsection begins by describing the Auto-Regressive (AR) and Moving Average (MA) models, before describing the Auto-Regressive Integrated Moving Average (ARIMA) method.

1. The AR model [7] extracts the influence of the values of the previous periods from those of the current period. This model is developed using the following linear equation.

$$Y_t = c + \varphi_1 . Y_{t-1} + \ldots + \varphi_p . Y_{t-p} + e_t \quad (2)$$

The parameter p indicates the AR order in the model and represents the delayed time period of the dependent variable. The remaining parameters of the equation, φ which represents the AR coefficient, y, which is the observed value, e the deviation of the series at the current instant, c is a constant [7].

2. The MA model [7] extracts the influence of the error terms from the previous period in the current period. This model is developed using the following linear equation:

$$Y_t = c + e_t + \theta_1 . e_{t-1} - \ldots - \theta_q . e_{t-q} \quad (3)$$

The parameter q indicates the MA order in the model and represents the delayed forecast errors. The remaining parameters of the equation, θ which represents the MA coefficient, y, the observed value, e the deviation of the series at the current instant, c is a constant [7].

3. The non-seasonal ARIMA model [6] [7] consists of three components, AR, Integrated (I) and MA, each component represented by a positive integer parameter, p, d and q respectively. These three components are combined in the following linear equation:

$$Y_t = c + \varphi_1 . Y_{d\ t-1} + \ldots + \varphi_p . Y_{d\ t-p} + e_t + \theta_1 . e_{t-1} - \ldots - \theta_q . e_{t-q} \quad (4)$$

I component [6], represented by parameter d and indicating the number of times the series has been differentiated to approximate stationary in time; AR component [4] [6], represented by parameter p means the delayed time period, estimated by Autocorrelation Function (ACF); MA component [4] [6] represented by parameter q indicates the order of the MA component and represents the delayed forecast errors, estimated by the Partial Autocorrelation Function (PACF).

It should be noted that parameters p and q are determined when the respective functions, ACF and PACF, cross the upper confidence interval for the first time. The confidence interval of the two functions is calculated as $\pm 1.96 / \sqrt{n}$, where the variable n, corresponds to the size of the historical data [6] [7]. Finally, seasonal ARIMA [7] extends the previous model, combining its components along with the seasonal component.

Time series forecasting with Prophet is more automated, due to its ability to find automatically inflection points in the data originated by changes in trend. A novelty of this method in relation to the previous one is the possibility of accommodating the existence of seasonal festive periods. The method combines three components, the trend, the seasonality and the festive periods, each modelled by some function [8]:

$$y(t) = g(t) + s(t) + h(t) + \epsilon t \quad (5)$$

The trend component, g(t), is modelled by a logistic function. The seasonality component, s(t), by a Fourier series. The festive periods, h(t), are adjusted by parameterization in the model. Finally, the error term, εt, represents the changes originated by circumstances that are not accommodated by the model [8]. Further information on the formulation details of each of these components can be found in [8].

4 Manual Parameterization of the Time Series

Figure 1 shows real data from a Telecom company graphically. It includes the variation of the number of recharges (used in telecommunications services), and the volume of internet data consumed (data) over the period from 2014 to 2017. There is a tendency of recharge decrease and data consumption increase. Seasonality exists there as temporary increases and decreases in certain months of each year. In the series of recharges, there is a lower consumption in the months of February, October and November, and a higher consumption in the months of January, August and December. First 3 years were used for analysis and training, the fourth year for prediction testing.

Figure 2 shows the stationary test of the time series of recharges using 2 methods. The figure above shows graphical visualization of moving average and deviation. The second, shows Dickey-Fuller statistical test presented below the figure.

Fig. 1. Consumption variation over 4 years

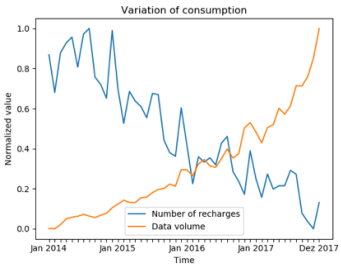
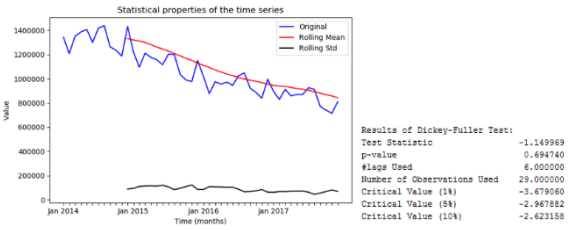
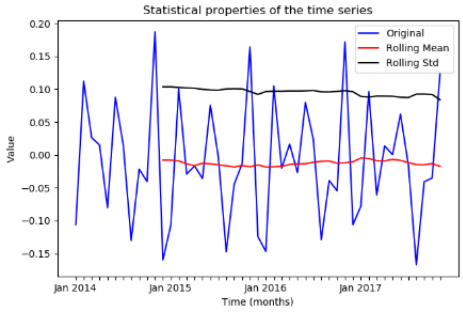


Fig. 2. Stationarity test of time series



We can see a decrease of the average over time. This variation over time indicates that the series is non-stationary. The result of the statistical test has a value greater than the critical value with a confidence level of 95% (-1.149969> -2.967882), also indicating that the series is non-stationary in time. Then a first-order differentiation transformation was applied to make the time series stationary in time. Figure 3 presents the stationary test of the series after this transformation.

Fig. 3. Stationarity test of the transformed time series

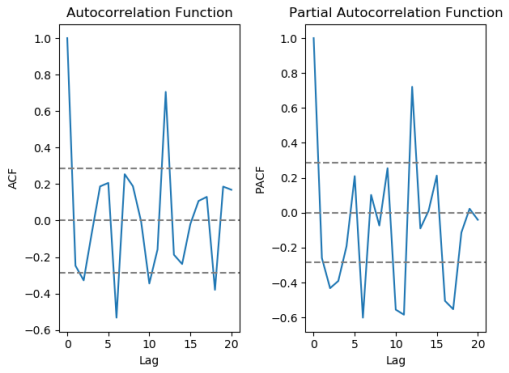



```
Results of Dickey-Fuller Test:
Test Statistic                  -4.815318
p-value                          0.000051
#lags Used                      10.000000
Number of Observations Used     24.000000
Critical Value (1%)             -3.737709
Critical Value (5%)             -2.992216
Critical Value (10%)            -2.635747
```

From the analysis of the previous graph, it is verified that the statistical properties of the time series have become approximately constant over time. The result of the statistical test is also less than the critical value for a 95% confidence level (-4.815318 <-2.992216), also indicating that the time series approached stationarity. Given that a differentiation was required to approximate the series of stationarity, then the parameter d of the model is equal to one.

The next step is to determine the values of the parameters p (order of the AR component of the model) and q (order of the MA component of the model) of ARIMA. To estimate these two parameters, we used the functions Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Figure 4 shows the graphs of these two functions.

Fig. 4. Graphs ACF and PACF



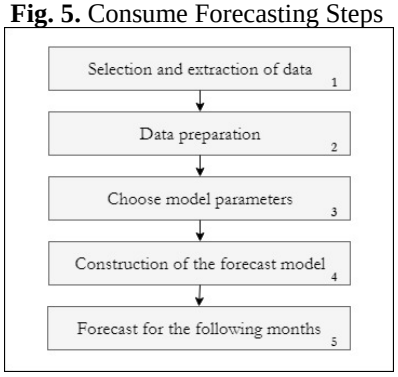
The graph on the left side of Figure 4, representing the ACF function, allows estimation of the value of parameter p, by looking at the position where the function crosses the confidence interval for the first time. In the graph this happens between p = 0 and p = 1. The graph on the right side is the PACF function, which allows to obtain the value of parameter q, also by crossing the function with the confidence interval. In the graph above, it happens between q = 0 and q = 1. In the experimental section we will compare the performance achieved by the forecasting model for this dataset using each of four possible parameter combinations (p = 0, d = 1, q = 0), (p = 1, d = 1, q = 0), (p = 0, d = 1, q = 1) and (p = 1, d = 1, q = 1).

Note that Prophet is much simpler to use than ARIMA, because of its ability to automatically find inflection points in the data, that is, points where trend changes [6].

5 Automation of the ARIMA model using exhaustive search

Automatic parameterization of the seasonal ARIMA model is important to allow its integration in a tool that can be used by managers without requirement of manual configuration or even knowledge of the details. It also accommodates changes in the behavior of the series. The proposed approach follows the steps shown in Figure 5.

Fig. 5. Consume Forecasting Steps



1. Selection of data, extraction and aggregation by month of a three-year historical period, related to the data that is intended to be forecasted.

2. Data preparation, subdivided into three tasks. The first converts the original date format to the format required by the template. The second constructs the input data structure in two-column model with date and the data consumption. The third applies a logarithmic transformation required by the method, which allows to attenuate the trend of the time series [6].

3. Choice of model parameters: This step aims to test and evaluate several models in order to choose the one that minimizes the forecast error. In this step, all combinations of values for parameters p, q and d are generated. The first two parameters can be zero, one or two, and parameter d can be zero or one [6]. This defined range of values takes into account manual parameterization tests as the ones presented in our experimental results, and also a certain flexibility to adapt to future data. The seasonality parameter is defined as 12 (yearly), corresponding to the seasonal period. After generating the combinations of parameters, each model instantiated is tested with historical values. The stationarity of the time series is a requirement for the application of the ARIMA method. Therefore, when the model is constructed with a non-stationary time series, the combination of parameters tested in this iteration is discarded, and the test advances to the next parameter combination.

4. Construction of the forecast model: after the choice of parameters, the forecast model is constructed with a recent three-year history, and with the parameters identified in the previous step.

5. Forecast of consumption of the following months: starts after the construction of the chosen model and consists of predicting the following months. The expected values are converted back to the scale of the original values by performing an exponential transformation (reverse operation of the logarithmic transformation).

6 Experimental Analysis

In the experiments presented in this section, we used real telecommunication data to compare the accuracy of seasonal ARIMA and Prophet methods when forecasting Telecom data, and to validate the parameterization mechanism of the proposed model.

6.1 Experimental Setup

The data used in this evaluation comes from a medium-sized telecommunications operator. Data was aggregated by month for a time period of four years. It consisted of consumption data, recharges (used in telecommunications services) and the volume of internet data consumed (data). Those are the same datasets already described in section IV. For both datasets we used the first three years for training and the fourth for testing the model.

The performance test (evaluation of execution time) was performed on a development machine with the following characteristics:

- Operating system: Windows 8 de 64 bits
- Processor: i5 de 2.50 GHz
- Memory RAM: 8.00 Gb
- Disk HDD: 297 Gb

6.2 Comparison between ARIMA and Prophet on Recharge dataset

Table 1 shows the RMSE and MAPE obtained using ARIMA with different combinations of parameter values, and Table 2 compares the forecasting errors of ARIMA and Prophet, choosing the best ARIMA result.

Tab. 1. Comparison of ARIMA errors on Recharges, 12 months

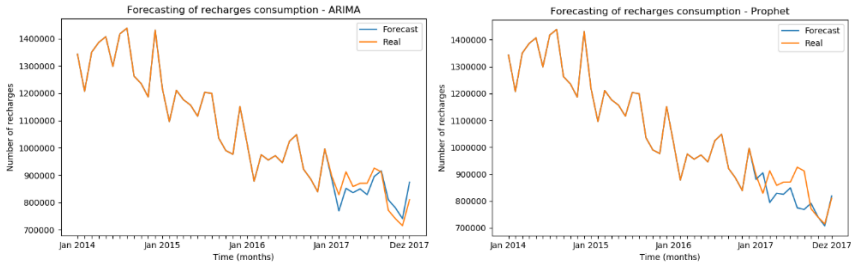
ARIMA (p, d, q)	RMSE	MAPE
(0, 1, 0)	4.08e+4	4.28
(1, 1, 0)	4.04e+4	4.24
(0, 1, 1)	3.99e+4	4.19
(1, 1, 1)	3.96e+4	4.14

Tab. 2. ARIMA versus Prophet on Recharges, 12 months

Method	RMSE	MAPE
ARIMA (1, 1, 1)	3.96e+4	4.14
Prophet	5.59e+4	6.3

From these results we can see that parameter configuration in ARIMA improves the error (RMSE or MAPE), and that ARIMA forecasting errors are much lower than Prophet for the Recharge dataset. Figures 6 show the forecast results graphically, comparing real to predicted values.

Fig. 6. Recharges Forecasting with ARIMA



From visual inspection of the results graphs we can see that both methods are able to deal effectively with both seasonality and trends. It is also clear that ARIMA outperforms Prophet for this dataset.

6.3 Comparison of various configurations

Table 3 presents the results of forecasting 12 months of internet data consumption using ARIMA and Prophet.

Tab. 3. ARIMA vs Prophet on Data Consumption, 12 months

Method	RMSE	MAPE
ARIMA	3.99e+13	8.50
Prophet	2.23e+13	9.88

Table 4 shows the comparison results for recharges and for data consumption (consume) over 3 months. In this case the training data was a whole year and the forecasting covered the next 3 months. Four forecasting runs were used, corresponding to the 4-year trimesters, the tables show the average and standard deviation of the errors over the four runs.

Tab. 4. ARIMA vs Prophet on Recharges and Data Consumption, 3 months

Method	RMSE		MAPE	
	Mean	Maximum	Mean	Maximum
ARIMA - Recharges	3.69e+4	4.97e+4	3.71	5.01
Prophet - Recharges	7.63e+4	1.05e+5	7.82	11.08
ARIMA - Data	2.59e+13	5.12e+13	6.99	17.0
Prophet - Data	4.67e+13	6.35e+13	8.11	13.26

We can see from the tables that Data consumption was slightly more challenging than Recharges for both methods (higher average error). ARIMA was always better

than Prophet (lower average error, almost half for both datasets), although ARIMA results have a higher standard deviation when compared with Prophet. Note that while for Recharges the sum of average error plus standard deviation or the maximum error were always smaller for ARIMA, in the case of Data Consumption one of the trimesters had the highest error for ARIMA, maximum error = 17% against 13% for Prophet, and the average error plus standard deviation is slightly higher for ARIMA. This was however an exception, as we could see by the results ARIMA was consistently better than Prophet for all other 12 and 3 months forecasting cases for both datasets.

6.4 Validation of automated ARIMA configuration

This section presents the validation results of the automatic parameterization mechanism of the ARIMA model described in section V. Table 5 shows the error of each combination tested by the exhaustive search approach regarding the 3-month and 12-months forecast, for both datasets tested. The exhaustive search parameter configuration approach obtains these errors for all cases and then chooses the one with lowest error, which is shown in bold. In all the tests it is verified that the implemented mechanism was able to automatically find the combination of parameters that minimizes the prediction error in the historical data.

The parameters chosen by the automatic approach for these 4 cases (Recharges and Data Consumption, 12 and 3 months forecasting) were always consistent with the established manual parameter configuration found (discussed and exemplified in sections III and IV). Note that the established manual approach involves iterative human inspection, running the Dickey-Fuller test or choosing visually, then successively differentiating and again testing using Dickey-Fuller or visual inspection until the decision thresholds are met. The exhaustive search simply replaces that tedious process with the automated version, with good results for these datasets and forecast objectives.

Another relevant issue is the runtime of the exhaustive search procedure, since it has to test a significant number of alternatives. We test the procedure runtime next.

Tab. 5. Automated parameter values finding by exhaustive search, 3 and 12 months

ARIMA (p, d, q)	Recharges (RMSE)		Data consumption (RMSE)	
	3 months	12 months	3 months	12 months
(0, 0, 0)	Discarded	Discarded	Discarded	Discarded
(0, 0, 1)	Discarded	Discarded	Discarded	Discarded
(0, 0, 2)	Discarded	Discarded	Discarded	Discarded
(0, 1, 0)	3.87e+4	4.08e+4	3.98+e13	5.67e+13
(0, 1, 1)	**3.69e+4**	3.99e+4	4.02+e13	4.82e+13
(0, 1, 2)	3.75e+4	4.03e+4	3.84+e13	5.89e+13
(1, 0, 0)	Discarded	Discarded	Discarded	Discarded
(1, 0, 1)	Discarded	Discarded	Discarded	Discarded
(1, 0, 2)	Discarded	Discarded	Discarded	Discarded

(1, 1, 0)	3.86e+4	4.04e+4	4.18e+13	5.13e+13
(1, 1, 1)	3.77e+4	**3.96e+4**	**2.59e+13**	**3.99e+13**
(1, 1, 2)	3.75e+4	4.03e+4	2.84e+13	5.55e+13
(2, 0, 0)	Discarded	Discarded	Discarded	Discarded
(2, 0, 1)	Discarded	Discarded	Discarded	Discarded
(2, 0, 2)	Discarded	Discarded	Discarded	Discarded
(2, 1, 0)	3.76e+4	4.06e+4	3.94e+13	6.48+e13
(2, 1, 1)	3.88e+4	3.97e+4	3.55e+13	7.36+e13
(2, 1, 2)	6.58e+4	4.01e+4	3.96e+13	1.19e+14

Table 6 shows the runtime of the automatic procedure (testing a significant set of parameter values alternatives) and compares it to the runtime of a single test as a ground truth. Note that the single test run is only used here for comparative reference, because running a single test automatically is not sufficient to configure the parameters, it must be done by a user inspecting the result, and most of the times, as happened for the datasets tested in this experimental section, will require further iterations of differentiating and running the test again.

Tab. 6. Runtime of parameter configuration by automated exhaustive search

Approach	**Time (s)**	
	Mean	**Maximum**
Recharges Forecast (automated)	6.38 ± 0.11	6.57
Data consumption (automated)	6.21 ± 0.39	6.92
Single test	0.30 ± 0.16	0.55
Single test	0.61 ± 0.03	0.66

These results show that the exhaustive search for parameter configuration takes about 6 secs, which is perfectly acceptable for the practical purposes of the decision support tool we were developing. The user has to wait for only 6 secs before the forecasting model does the forecast, since it is searching for the correct parameters. Nevertheless, we note that, although the procedure works fine for the sizes of datasets that our tool works with, it is important to develop improved parameter finding approaches in the future, to be able to handle much bigger datasets. Therefore, we identify as future work the possibility of improving the automated parameter configuration procedure. A simple way to scale to large datasets would involve sampling, to reduce the dataset to a size that is tractable by exhaustive search, the other alternative would be to apply heuristics to reduce the search space, and a third alternative would involve both. We reserve this study for future work.

7 Conclusion and future work

In this work we studied the application of time series forecasting methods ARIMA and Prophet to real Telecom data, with the aim of integrating the best performing one in a practical tool for decision support. The two methods were presented along with

their parameterization processes. These two methods were compared in the forecast of consumption using real telecommunications data. Since the ARIMA model requires a great deal of knowledge in its parameterization, an automatic parameterization procedure was proposed and validated. This allows the approach to do every step of the data forecasting pipeline automatically. This way the approach was integrated into a tool used by managers to view the forecasts without requiring any knowledge of the data preparation and parameterization process for ARIMA forecasting.

This work also showed that, despite the great variation of consumption during the year, due to the existence of seasonality and trends, it is possible to make approximate forecasts with consumption data in the telecommunications area. From the results obtained and the comparison of the two time series methods, it was possible to obtain a minimum MAPE of 3.71% in the three-month forecast and of 4.14% in the twelve-month forecast. Of the two methods tested, the ARIMA model presented better prediction results in relation to Prophet.

For future work we intend to develop models of the Holt-Winters time-series method and compare with the methods studied in this work. In addition, we intend to investigate alternatives that improve the exhaustive search method in the automatic parameterization of ARIMA, using heuristics.

References

[1] StatsModels Statistics in Python, available in http://www.statsmodels.org/devel/generated/statsmodels.tsa.arima_model.ARIMA.html, consulted in 2018-03-20.

[2] Prophet Forecasting at scale, available in https://facebook.github.io/prophet/, consulted in 2018-04-16.

[3] BRANCO, S. T., & DE SAMPAIO, R. J. B. A New Artificial Neural Networks Forecast Model in Telecommunications, consulted in 2017-10-28.

[4] Wang, M., Wang, Y., Wang, X., & Wei, Z. (2015). Forecast and Analyze the Telecom Income based on ARIMA Model. The Open Cybernetics & Systemics Journal, 9(1), consulted in 2018-03-15.

[5] Is Prophet Really Better than ARIMA for Forecasting Time Series Data, available in https://blog.exploratory.io/is-prophet-better-than-arima-for-forecasting-time-series-fa9ae08a5851, consulted in 2018-05-16.

[6] A comprehensive beginner's guide to create a Time Series Forecast, available in https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/, consulted in 2018-03-15.

[7] MATH6011: Forecasting, available in https://www.southampton.ac.uk /~abz1e14/papers/Forecasting.pdf, consulted on 2018-03-22.

[8] Taylor, S. J., & Letham, B. (2017). Forecasting at scale. The American Statistician, (just-accepted), consulted in 2018-04-16.

[9] MAE and RMSE - Which Metric is Better, available in https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d, consulted in2018-03-20.

Multivariate forecasting of extreme wave climate and storm evolution

Andrea Lira-Loarca[1[0000-0003-2251-6684]], Manuel Cobos[1[0000-0002-5896-3630]], Asunción Baquerizo[1[0000-0002-3339-0653]] and Miguel A. Losada[1[0000-0002-2071-3134]]

[1]Andalusian Institute for Earth System Research. University of Granada. Av. del Mediterráneo s/n, 18006 Granada, Spain
aliraloarca@ugr.es, mcobosb@ugr.es, abaqueri@ugr.es, mlosada@ugr.es

Abstract. The design and management of a coastal structure should be addressed as an optimization problem that depends on random multivariate climate variables and takes into account the damage progression along its useful life. An optimal design requires an accurate characterization and evolution of extreme events. In this work, we present a simple and efficient methodology to simulate time series of storm events including several maritime variables. It includes the use of advanced statistical techniques such as non-stationary mixture distribution functions, a vector autoregressive model and a copula model for the characterization of extreme maritime variables and storm evolution. The proposed methodology is able to efficiently provide a time series of simulated storms which reproduces the co-dependence between variables and the temporal distribution of the historical time series.

Keywords: Forecasting Maritime Climate, Storm Evolution, VAR, Copula Model

1 Introduction

In recent years, there has been an increased concern regarding climate change impacts in coastal areas. According to the Intergovernmental Panel on Climate Change Fifth Assessment Report [1], coastal systems and low-lying areas will increasingly experience adverse impacts such as submergence, coastal flooding, and coastal erosion due to sea level rise, increase in the frequency and intensity of storms and the associated storm surges.

Without adaptive measures, almost a million people per year are expected to be affected by coastal flooding in a climate change scenario and the economic losses and costs due to flooding, salinity intrusion, land erosion and migration are projected to be about 17 billion dollars by 2100 in Europe [2].

The traditional design practice of coastal structures and protective measures is based on the statistical analysis of isolated climate variables such as the significant wave height by means of its return period without taking into account the dependence be-

tween different forcing agents nor their temporal variation. Likewise, storm characterization and evolution is done using simple geometric shapes like the Equivalent Triangular Storm [3-4] with no information given on the evolution of the variables on the sea states that define the storms.

Nevertheless, most of the climate variables have the same origin and should not be treated as independent. Therefore, univariate return periods are not sufficient to estimate a storm and its evolution as it is a multivariate problem depending on the significant wave height (H_s), mean wave period (T), wave direction (θ), storm duration (D) and interarrival time (I), among others.

Accordingly, there is a new dimension added to the traditional design practice in maritime engineering with a need for more advanced statistical techniques for extreme analysis that take into account the complex interactions and predictions of maritime climate variables.

This work presents a simple and efficient methodology to simulate time series of storm events including several maritime variables. This methodology includes the use of advanced statistical techniques such as non-stationary parametric distributions [5] to marginally characterize each variable, a vector autoregressive (VAR) model to describe the temporal dependence between variables, and a copula model to link the seasonal dependency of the storm duration and the interarrival time between consecutive storms.

2 Methodology

The methodology comprises the following steps (Fig. 1):

a. Storm definition: For a given threshold value $H_{s,u}$, the storm events (duration, D) and the succeeding calm periods (interarrival times, I) are identified over the historical time series of the significant wave height, H_s. The concomitant mean periods, T_p, and mean wave directions, θ_m , are also obtained.
   Minimum values of D and I between consecutive events are set to ensure independency between events.
b. H_s, T_p and θ_m are characterized using non-stationary marginal mixed distribution functions.
c. The dependence between the duration D and interarrival times I, is analyzed seasonally using a copula model.
d. A vector autoregressive model is used to analyze the temporal dependence and interdependence of the variables and therefore characterize the storm evolution.
e. A continuous time-series of maritime variables is obtained using the VAR model and the fitted distribution functions. A new threshold, $H_{s,u}'$, is set in order to define independent storms, maintaining the similarity between the CDFs of the storms' duration D.
f. Storms' time-series simulation: The simulated independent storms are then distributed throughout time by simulating the interarrival time from the copula model of I conditioned to the corresponding duration, D.

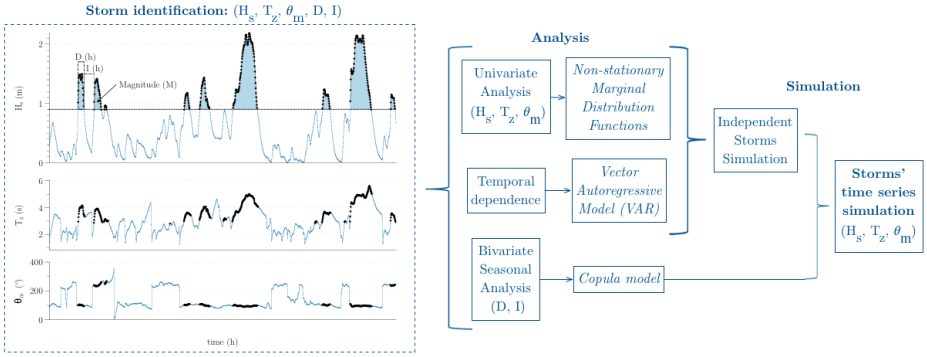


Fig. 1. Diagram of the methodological framework

3 Results

A 50-year time-series of simulated storms in Granada, Spain was done using historical wave climate from the point SIMAR-2041080 (Puertos del Estado) with hourly data from 1958-2018. H_s and T_p were fitted using Exponential and Lognormal distributions, respectively. θ_m was fitted using a mixture distribution model of two truncated Normal distributions. A Clayton copula was used to study the dependence of D and I. Fig. **2** presents the scatter plots of the main variables of historical and simulated time series. It can be observed that the model reproduces fairly well the joint stochastic character of the random time series as well as the temporal dependence given that the simulated time series presents a similar monthly storm frequency as the historical data (Fig. 3). During the conference, we will present a more in-depth description of the proposed methodology and the results of its implementation in different climate maritime regions of the Spanish coast.

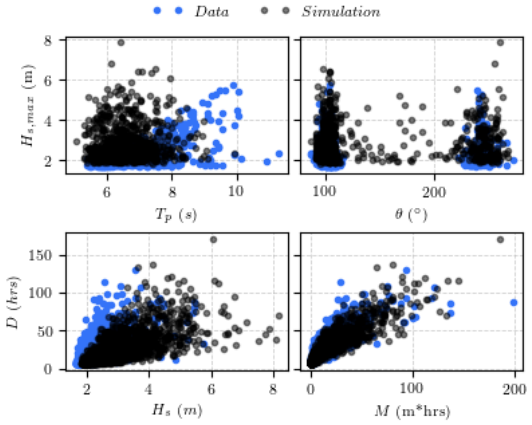


Fig. 2. Scatter plots of data and simulations

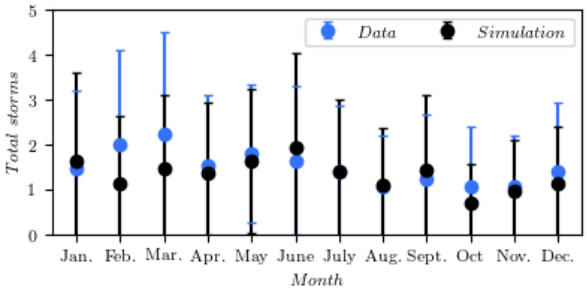


Fig. 3. Monthly average storm frequency

Acknowledgements

This research was supported by the project AQUACLEW and the research group TEP-209 (Junta de Andalucía). Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Comission [Grant 690462].

References

1. IPCC: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [C.B. Field, V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA (2014).
2. Hinkel, J., Nicholls, R.J., Vafeidis, A.T., Tol, R.S., Avagianou, T.: Assessing risk of and adaptation to sea-level rise in the European Union: an application of DIVA. Mitigation and Adaptation Strategies for Global Change 15(7), 703–719 (2010).
3. Boccotti, P: Wave Mechanics for Ocean Engineering. Elsevier Science, Oxford (2000)
4. ROM-1.0-09: Recommendations for the Project Design and Construction of Breakwaters (Part I: Calculation and Project Factors. Climatic Agents). Puertos del Estado (2009).

Pattern similarity-based load forecasting applied to unit commitment problem

Guilherme Costa Silva[1], Adriano C. Lisboa[2], Douglas A. G. Vieira[3], and Rodney R. Saldanha[1]

[1] Graduate Program in Electrical Engineering - Universidade Federal de Minas Gerais - Av. Antônio Carlos, 6627 31270-901, Belo Horizonte, MG, Brazil
guicosta@ufmg.br,rodney@cpdee.ufmg.br

[2] Gaia Solutions on Demand - Rua Professor José Vieira Mendonça, 770, 31270-901, Belo Horizonte, MG, Brasil
adriano.lisboa@gaiasd.com

[3] ENACOM Handcrafted Technologies - Alameda das Latânias, 215, 31270-800, Belo Horizonte, MG, Brasil
douglas.vieira@enacom.com.br

Abstract. The short-term load forecasting problem consists of estimating the future demand for a period of up to one week and has fundamental importance in power system studies. This problem can also be modeled in terms of pattern similarity, allowing the use of regression algorithms based on machine learning applied to time series analysis. In this work, three neural network-based models are evaluated and forecasted data is applied to the unit commitment problem, providing a schedule for power plant operations in some case studies. These experiments present interesting results, proving the applicability of the models and the combined strategy.

Keywords: Pattern similarity, Short term load forecasting, Unit commitment problem, Power systems

1 Introduction

The load demand can be considered as one of the most important information to be used in the operation of energy systems. Operations in power systems are very dependent on these information, which results in the successful operation of generators, substations and lines, as well as the planning of new units in these systems.

A large number of models for load forecasting have been proposed, some of them based on regression, statistical analysis, smoothing, or parametric analysis, some examples are integrated auto-regressive models based on moving averages (ARIMA) and exponential smoothing models of Holt-Winters according to [1]. Finally, research in computational intelligence have grown and models based on neural networks became widely used as in [2–4].

In [5], a load prediction model based on pattern similarity is proposed. This model facilitates the use of several machine learning algorithms in time series

based problems and provides intuition prediction based on previously known patterns.

Thus, the realization of the short-term load forecasting (STLF) plays a fundamental role, especially in the control, scheduling and safety tasks of the electrical systems. Among them, the Unit Commitment problem (UCP), considered in this work as in [6], is a recurrent optimization task in power systems which involves scheduling the generating units status (ON/OFF) in order to meet the load demand with minimum generation cost satisfying the different existing constraints. UCP tasks are highly dependent on a successful load forecasting so that generators may satisfy the demand with less costs of operation.

The work in question discusses some algorithms based on the pattern similarity model and their application in STLF problems, applying models based on self organizing maps (SOM) neural networks proposed in [4] and employing their application in UCP considering some adaptation of load forecast case studies to the problem.

2 Basic review

2.1 Short term load forecasting

The problem of short-term load forecasting is to predict future demand in a range of a few hours before up to a few weeks ahead, as shown in Figure 1. In this problem, it is possible to predict statistically a distinct pattern of consumption resulting from the totality of individual loads, according to [7].

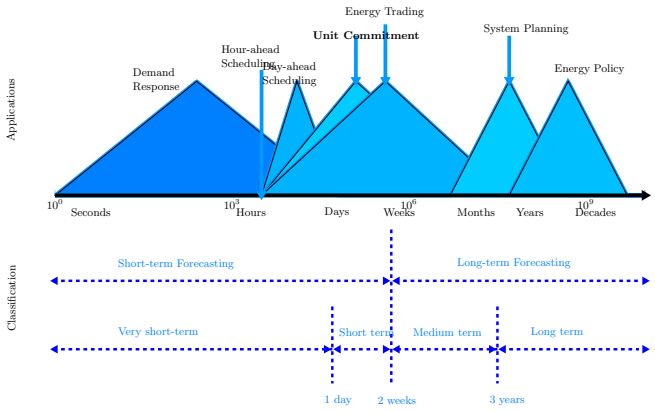


Fig. 1. Taxonomy and applications of Load Forecasting problems, according to [8].

Some classes of factors may influence the behavior of the system load, such as seasonal, economic and climatic factors, or even some random effects. In [9], three

objectives for the problem are listed, such as scheduling generation, providing safer and more reliable operation of the plants, and increasing system reliability.

In this way, it is possible to model the load demand in time series considering some tendencies presented in the behavior of the operations of energy systems. Thus, such information can be predicted, for example in Box and Jenkins time-series models.

The methodology studied in this work is the approach proposed in [5], which treats the prediction of load using neural network methods based on pattern similarity applied to load forecasting according to the daily demand considered. Result of the adopted techniques are applied to Unit Commitment tasks.

2.2 Pattern similarity-based load forecasting approach

In [10], main aspects of the load forecasting problem were discussed, as well as the adequacy of pattern recognition methods to the problem. Considering some factors a seasonality and trends, it is possible to transcribe a simple and univariate problem based on similarity metrics, based on a simple optimization process that requires few procedures such as defining the sampling period.

Short-term load forecasting problem can be defined by sampling energy demand in days, with its hourly information as a dimension of the problem, representing a fragment of the time series, used in pairs x, which represents the historical information to be analyzed, and y, which represents the information to be forecasted.

The prediction model based on pattern similarity can be defined by the following steps:

1. Mapping of the time series (seasonal cycles) in paired patterns $\mathbf{x}$ and $\mathbf{y}$.
2. Selection of similar patterns $\mathbf{x}$ to the query pattern $\mathbf{x}^{\star}$.
3. Definition of the pattern $\mathbf{y}$ that pairs with $\mathbf{x}$ in order to forecast $\hat{\mathbf{y}}$ related to $\mathbf{x}^{\star}$.
4. Decoding of $\hat{\mathbf{y}}$ in the seasonal cycle referring to $\hat{\mathbf{z}}$.

This prediction model considers the time series pre-processing to treat specificities in the series to be analyzed or simply to perform normalizations in the data, according to the following equations:

$$x_t = f_x(z_{i,t}, \phi_i) \tag{1}$$

$$y_t = f_y(z_{i+\tau,t}, \phi_i) \tag{2}$$

where z is the original time series, x is the analyzed pattern and y is the pattern to be forecasted, i refers to the period analyzed before forecast, t is the hour of the day, τ is the period to be forecasted related to i and ϕ_i is the decoding variable, usually the daily average load. Usually, pre-processing is defined by the following functions:

$$f_x(z_{i,t}, \phi_i) = \frac{z_{i,t} - \phi_i}{\sqrt{\sum_{j=1}^{N} z_{i,j} - \phi_i}} \tag{3}$$

$$f_y(z_{i+\tau,t}, \phi_i) = \frac{z_{i+\tau,t} - \phi_i}{\sqrt{\sum_{j=1}^{N} z_{i,j} - \phi_i}} \tag{4}$$

Note that in the construction of **y** only prior data used, once **y** values are not known at the time of the forecast.

In this model, similarity functions are defined, usually based on linear or non-linear mapping of distance metric D in order to provide comparisons among **x** patterns. In training, data related to **y** patterns are also analyzed and the correlation between measured distances is verified. In the test, the **y** pattern is estimated by the **x**-pattern observations.

Usually, the metrics used to assess forecasts quality is the mean absolute percent error (MAPE) between the forecasted data F_t and the actual data A_t in the period of n instants, as follows:

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \frac{|A_t - F_t|}{A_t} \tag{5}$$

This metric is commonly used in load forecasting problems. Its interquartiles, applied in [10], indicate information on error scattering and smaller values of them point to more accurate models, since errors are centered around the average value.

With this, supervised algorithms can be used so that it is able to map $x \rightarrow y$ and then new **x** patterns are presented and **y** patterns are obtained as outputs.

2.3 The unit commitment problem

Basically, the UCP [11] deals with the scheduling of generators according to their status (ON or OFF) during a given period (usually defined by hours, per day or week) in order to supply demand at the given hours. The problem is defined by its cost objective and some constraints, which can vary according to some formulations.

The adopted formulation is based on [11–13]. The cost objective is defined as in (6), with its components and variables defined in (7) to (11):

$$\min Cost = \sum_{t=1}^{T} [Cp_t + Ce_t + Su_t + Sd_t] \tag{6}$$

$$Cp_t = \sum_{n=1}^{N_{Gen}} a_n G_{n,t}^2 + b_n G_{n,t} + c_n \tag{7}$$

$$Ce_t = \sum_{n=1}^{N_{Gen}} \alpha_n G_{n,t}^2 + \beta_n G_{n,t} + \gamma_n \tag{8}$$

$$Su_t = \sum_{n=1}^{N_{Gen}} \min(0, S_{n,t} - S_{n,t-1} + 1) \times Csu_n \tag{9}$$

$$CSu_t = \begin{cases} CS_{hot,n} & \text{if } S_{n,t} \leq ct_n \\ CS_{cold,n} & \text{otherwise} \end{cases} \tag{10}$$

$$Sd_t = \sum_{n=1}^{N_{Gen}} \max(0, S_{n,t} - S_{n,t-1} - 1) \times -Csd_n \tag{11}$$

where:

Cp_t	total production costs at hour t, in $\$/MW$;
Ce_t	total emission costs at hour t, in $\$/ton$;
Su_t	total unit startup cost at hour t, in \$;
Sd_t	total unit shutdown cost at hour t, in \$;
$G_{n,t}$	active power of unit n at hour t,(MW);
a_n, b_n, c_n	production cost coefficients of unit n;
$\alpha_n, \beta_n, \gamma_n$	emission cost coefficients of unit n;
$S_{n,t}$	state of unit n at hour t,[0,1];
Csu_n	startup costs of unit n,\$;
$Cs_{hot,n}$	hot startup costs of unit n,\$;
$Cs_{cold,n}$	cold startup costs of unit n,\$;
ct_n	cold start-up threshold for generation n,\$;
Csu_n	shutdown costs of unit n,\$;

The UCP equality and inequality constraints are represented as follows:

$$\forall t \sum_{n=1}^{N_{Gen}} (G_{n,t} \times S_{n,t}) = \sum_{l=1}^{N_{Load}} P_{l,t} \tag{12}$$

$$G_{min,n} \leq G_{n,t} \leq G_{max,n} \iff S_{n,t}, \forall n, t \tag{13}$$

$$-Rp_{down,n} \leq G_{n,t} - G_{n,t-1} \leq Rp_{up,n} \iff S_{n,t} - S_{n,t-1} = 0, \forall n, t \tag{14}$$

$$\forall t \sum_{n=1}^{N_{Gen}} (G_{n,t} \times S_{n,t}) = \sum_{l=1}^{N_{Load}} P_{l,t} + Sr_t \tag{15}$$

$$\forall t, n \sum_{h=t-T_{ON,n}}^{t-1} (S_{n,h} = T_{ON,n}) \iff S_{n,t} - S_{n,t-1} = -1 \tag{16}$$

$$\forall t, n \sum_{h=t-T_{OFF,n}}^{t-1} (S_{n,h} = 0) \iff S_{n,t} - S_{n,t-1} = +1 \quad (17)$$

where:

$P_{l,t}$	amount of load demand at t time interval and l load point, in MW;
$G_{min,n}$	minimum power limits of generator n, in MW;
$G_{max,n}$	maximum power limits of generator n, in MW;
$Rp_{up,n}$	ramp rate upper limit of generator n;
$Rp_{down,n}$	ramp rate lower limit of generator n;
Sr_t	forecasted spinning reserve at t hour;
$T_{ON,n}$	minimum up time of generator n;
$T_{OFF,n}$	minimum down time of generator n;

The UCP can be solved by different optimization methods, such as Dynamic Programming [12, 6, 13], and some meta-heuristic methods as well [14, 15]. Regardless of the method, this problem is dependent on a successful load forecasting task.

An earlier work in [16] has performed some tests considering hourly and daily forecasting by applying a feed-forward neural network to forecast step, the result was applied to unit commitment schedule decisions. Although the concept of pattern similarity has not been developed at all, the methodology is directly linked to this concept.

This work will apply some load forecasting methods based on pattern similarity in order to provide information for a dynamic programming-based unit commitment method in some benchmark cases. Both forecasting and generation scheduling will be evaluated, as well as the overall process performed in the research.

3 Experiments

The experiments performed in this work consider a hypothetical system in which the demand from New England area is supplied by generators from IEEE benchmark systems. There are five phases, two of them for data analysis and the others related to load forecasting, unit commitment and the overall analyses, as illustrated in Figure 2 and described throughout this work.

3.1 ISO New England demand

The ISO New England dataset, available in [17], consists of a database containing information regarding the 8 regions within New England area (three of them in Massachusetts state, see Figure 3), including the measured demand, and the corresponding prices. These data also served as a basis for a complete and complex database proposed in [18], also developed for UCP tests.

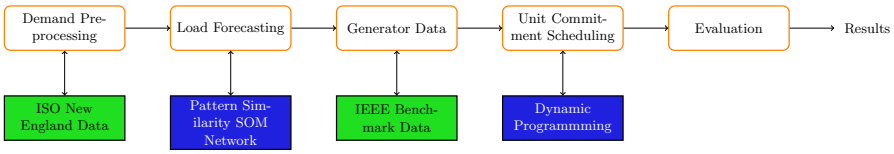


Fig. 2. Flowchart of the proposed experiments, described by each task, databases (in green/black), and methods used (in blue/white).

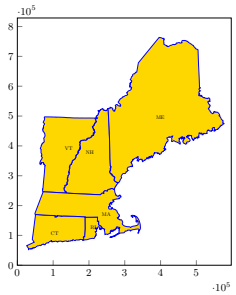


Fig. 3. New England region and its states.

This dataset is used during the STLF phase, in which the pattern similarity-based method is applied to one of 8 zones, The training data consists of load demand between 2007 and 2012, which is processed in order to forecast information regarding 2013 demand.

3.2 IEEE benchmark systems

Two benchmarks systems from IEEE, namely IEEE 14-bus and 30-bus, all of them illustrated in Figure 4, are tested during the unit commitment phase and their generators are scheduled in order to handle the demand forecasted during STLF phase.

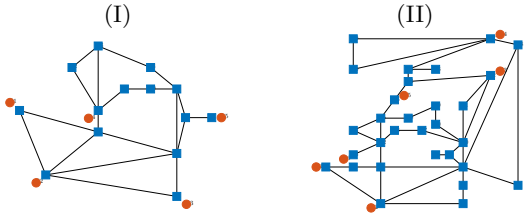


Fig. 4. Illustration of IEEE benchmark systems considered in this work: (I) 14-bus system, (II) 30-bus system

All database used for experiments are given in [6], except for the demand, which is replaced by the ISO NE data, as previously presented.

3.3 Pattern similarity SOM Network

In order to provide information regarding load demand, four different load forecast methods are employed and tested in this research. Three of these methods are based on Self-organizing maps neural network developed according to the pattern similarity principles studied in [4] as further explained.

Algorithm 1. Self-organizing map model 1.

1. Concatenation of the paired $\mathbf{x}$ and $\mathbf{y}$ patterns in the pattern $\mathbf{z}$;
2. Train SOM network according to patterns $\mathbf{z}$;
3. Presentation of the query pattern $\mathbf{x}^\star$ and assigning it to the nearest neuron w_j according to $j^\star = \text{argmin}\|x^\star - w_{x,j}\|$;
4. Reconstruction of the $\mathbf{y}$ pattern paired with the query pattern based on the $\mathbf{y}$ part of the neuron $w_{y,j}$.

Algorithm 2. Self-organizing map model 2

1. Independent grouping of patterns $\mathbf{x}$ and $\mathbf{y}$, with two SOMs for each pattern;
2. Estimation of the conditional probabilities $P(C_{y,l}|C_{x,k})$.
3. Presentation of the query pattern $\mathbf{x}^\star$ and assigning it to a set of nearest neurons $w_{x,j}$;
4. Reconstruction of the $\mathbf{y}$ pattern paired with the query pattern based on neuron $w_{y,k}$ from SOM of pattern $\mathbf{y}$ and probabilities $P(C_{y,l}|C_{x,k})$.

Algorithm 3. Self-organizing map model 3

1. Only $\mathbf{y}$ patterns are developed;
2. neurons w_j are labeled according to the following information: day numbers i and day types δ;
3. Calculation of the number of entries e_j which satisfy two conditions:
   (a) query pattern is from the same period of the year as the forecasted $\mathbf{y}$ pattern;
   (b) day type is the same as for the forecasted $\mathbf{y}$ pattern.
4. Reconstruction of the $\mathbf{y}$ pattern paired with the query pattern based on neuron w_j and label entry numbers e_j.

These three methods are tested with their performance mutually compared. Forecasting results are used in unit commitment phase.

3.4 Dynamic programming

For solving UCP, a Dynamic Programming (DP) approach proposed in [6] is used, this approach defines a mechanism named hourly state restriction, in which secondary states are defined, but there are some of these states omitted in favor of feasibility, the Figure 5 illustrates how it works.

In this work, DP evaluation of UDP is performed in two approaches as follows:

– Daily, with the 24 hours of a day evaluated, considered by its type:

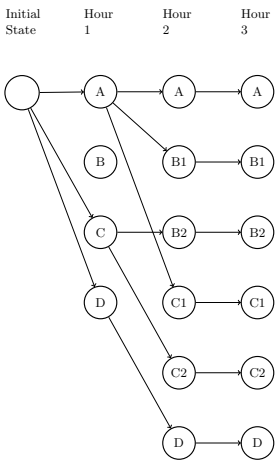


Fig. 5. Description of the Dynamic Programming approach used in this work, as in [6].

1. Working days;
2. Saturdays;
3. Sundays;
4. Holidays.

– Weekly, with the hours of a week, starting by monday, being evaluated at once and considered by its season:

1. Winter;
2. Spring;
3. Summer;
4. Autumn.

These data are obtained considering the mean of historical hourly data defined by the described period.

3.5 Solutions evaluation

After STLF and UDP phases, the UDP solutions obtained during these tests are evaluated according to a hypothesis test, which is defined as follows:

$$\Omega_S = \mathrm{H}(S_{g,t}, S^{\star}_{g,t}) \tag{18}$$

where $S^{\star}_{g,t}$ is the ideal solution, which is obtained considering the actual load demand of test data. The χ^2 test can be used in this case, in order to measure their closeness according to the ideal case.

4 Results

4.1 Load Forecasting

Load forecasting results are presented in Table 1, according to the Control Area and the 8 zones.

Table 1. Load forecasting results per zone, with best solutions in bold.

Zone Dataset	SOM Model 1		SOM Model 2		SOM Model 3	
	MAPE	IQR	MAPE	IQR	MAPE	IQR
ISONE CA	3.616625	3.440646	**3.495636**	**3.313823**	4.131780	3.605780
ME	3.184038	2.698629	**2.743899**	**2.024705**	2.988718	2.488731
NH	4.022246	4.022246	**3.904023**	**3.281043**	4.139372	3.558926
VT	3.114933	3.531652	**3.105123**	**2.655316**	3.201551	2.740366
CT	4.936994	4.519499	4.391825	3.490205	**4.208487**	**4.255713**
RI	4.649350	4.005369	**3.970818**	**3.388785**	4.287126	4.059443
SEMASS	4.536004	4.144500	**4.117994**	**3.353949**	4.590180	4.008678
WCMASS	3.890134	3.983390	**3.634790**	**3.448429**	4.268935	3.403963
NEMASSBOST	4.555781	4.592830	**3.790459**	**3.526698**	4.750288	4.221285

For these tests, the second model has presented results more closer to the actual data, except for the Connecticut zone, in which third model has presented a better result. In fact, the performance was similar, which implies that results in next phase may not present relevant differences.

4.2 Unit Commitment

In this phase, we consider the demand from Zone 6 (Rhode Island), and its forecasting will be applied to the IEEE benchmark systems through DP. For compatibility sake, the per-unit system was applied to generators and loads, demand used by the evaluated system is about 25% and 30% of the total demand forecasted for IEEE 14 and 30 bus respectively.

Results are presented in Figure 6 for IEEE 14-bus system and in Figure 7 for IEEE 30-bus system, for both daily and weekly approaches.

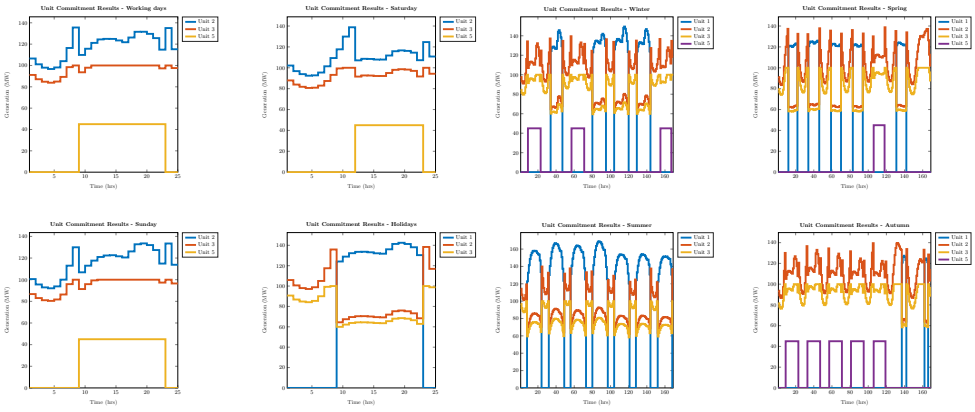


Fig. 6. Unit Commitment results for IEEE 14-bus System.

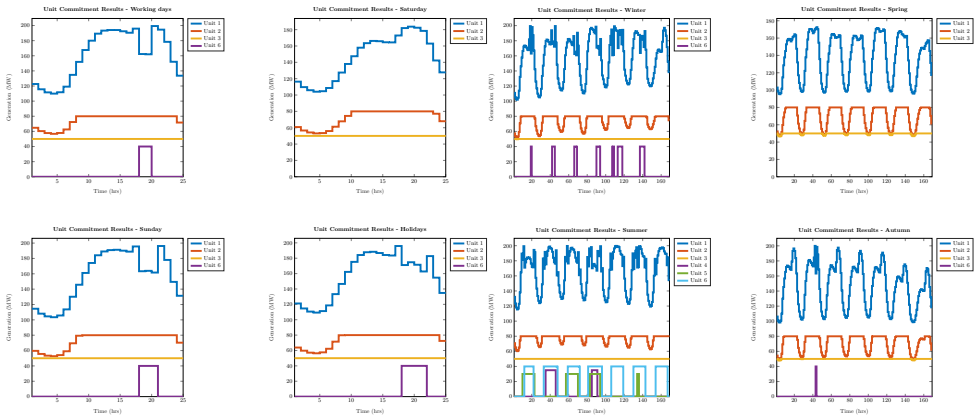


Fig. 7. Unit Commitment results for IEEE 30-bus System.

Solutions provided by DP in daily and weekly UC problems imply in different scheduling according to type of day and the season of year as well.

4.3 Overall Evaluation

In the final evaluation, observed planning is evaluated according to the expected values. As calculated through Ω_S, for all planning results, null hypothesis was rejected, pointing that planning was according to expected, with p-values of 1 and $\chi^2 = 0$, showing that the unit commitment strategy may correspond the actual demand.

5 Conclusions and future work

This work presented the application of pattern similarity to short-term load forecasting and unit commitment. The presented results confirm that good forecasts may reinforce strategies for generation schedules.

These preliminary results may be reinforce with the use of other forecasting techniques, such as Fuzzy Time Series, and other UC approaches, based on evolutionary algorithms, among other examples.

While the influence of load forecasting at UC problem is observed, another interesting study is the influence of both problems (mainly UC) in Generation Expansion Planning, which can be discussed with some scenarios in which new generators are introduced to the system, providing a complete analysis of the integration among these problems.

Acknowledgments. The authors acknowledge the support of CNPq, FAPEMIG and CAPES in this study. Special thanks to Vladimir Stanojevic for his DP implementation in MATLAB.

References

1. R. Weron, *Modeling and Forecasting Electricity Loads and Prices*, Wiley, Chichester, 2006.
2. K. Lang , M. Zhang, Y. Yuan, *Improved Neural Networks with Random Weights for Short-Term Load Forecasting*. PLOS ONE 10(12): e0143175, 2015.
3. S. Wei, L. Mohan, *Application of improved artificial neural networks in short-term power load forecasting* Journal of Renewable and Sustainable Energy, 7:4, 2015.
4. G. Dudek, *Neural networks for pattern-based short-term load forecasting: A comparative study*, Neurocomputing, vol. 205 (2016), pp. 64–74.
5. G. Dudek, *Pattern similarity-based methods for short-term load forecasting - Part 2: Models*, Appl. Soft Comput., vol. 36 (2015), Pages 422–441.
6. W. J. Hobbs, G. Hermon, S. Warner and G. B. Shelbe, *An enhanced dynamic programming approach for unit commitment*, in IEEE Transactions on Power Systems, vol. 3, no. 3, pp. 1201-1205, Aug. 1988.
7. G. Gross, F. D. Galiana, *Short-term load forecasting*, Proceedings of the IEEE , vol.75, no.12 (1987), pp.1558–1573.
8. T. Hong, S. Fan, Probabilistic electric load forecasting: A tutorial review, International Journal of Forecasting, Volume 32, Issue 3, 2016, pp. 914–938.
9. M. Q. Raza, A. Khosravi, *A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings*, Renewable and Sustainable Energy Reviews, vol. 50 (2015), pp. 1352–1372.
10. G. Dudek, *Pattern similarity-based methods for short-term load forecasting - Part 1: Principles*, Appl. Soft Comput., vol. 37 (2015), Pages 277–287.
11. M. F. Anjos, A. J. Conejo, Unit Commitment in Electric Energy Systems, Foundations and Trends in Electric Energy Systems, vol. 1, no. 4, 2017, pp. 220–310.
12. Chao-An Li, R. B. Johnson and A. J. Svoboda, *A new unit commitment method*, in IEEE Transactions on Power Systems, vol. 12, no. 1, pp. 113-119, Feb. 1997.
13. S. P. Karthikeyan, K. Palanisamy, C. Rani, I. J. Raglend, D. P. Kothari, *Security constrained unit commitment problem with operational, power flow and environmental constraints*, WSEAS Transactions on Power Systems Vol. 4 no.4 (2009), pp. 53–56.
14. A. S. Uyar and B. Turkay, *Evolutionary Algorithms for the Unit Commitment Problem*, Turkish Journal of Electrical Engineering, vol.16, no.3 (2008), pp. 239–255.
15. A. Azizivahed, S. Ghavidel, M. J. Ghadi, L. Li and J. Zhang, *A novel reliability oriented bi-objective unit commitment problem*, 2017 Australasian Universities Power Engineering Conference (AUPEC), Melbourne, VIC, 2017, pp. 1-6.
16. A. A. Girgis, S. Varadan, *Unit commitment using load forecasting based on artificial neural networks*, Electric Power Systems Research, vol. 32, no. 3 (1995), pp. 213–217.
17. ISO-NE Homepage [Online]. Available: http://www.iso-ne.com/
18. D. Krishnamurthy, W. Li and L. Tesfatsion, *An 8-Zone Test System Based on ISO New England Data: Development and Application*, in IEEE Transactions on Power Systems, vol. 31, no. 1, pp. 234-246, Jan. 2016.

Modified Granger Causality in Selected Neighborhoods

Martina Chvosteková

Institute of Measurement Science of Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava 4, Slovakia
martina.chvostekova@savba.sk,

Abstract. Although Granger causality is a widely used technique to detect causal relationship, its direct application for non-linearly modeled data is not appropriate. There have been proposed several extensions to non-linear cases, but none of the modifications identifies the correct causal connectivity in general. We present a new measure for evaluation of a causal effect between two time series, which is based on the selected local approximations of time delay embedding reconstruction of state space by a linear regression model. The novel causal measure reflects the proportion of variation of the modeled variable, fitted by a linear regression model including past of the modeled variable and of the second variable as predictors, actually predicted by the past of the second variable only. The suggested measure should be used to identify both linear and nonlinear causal effects. The proposed estimator is applied to six data sets with known causal connections and dynamics.

Keywords: Granger causality, time-delay embedding reconstruction, linear regression model, prediction error

1 Introduction

Assessing the presence of directional interactions between two time series of a multivariate data set is an important issue in diverse areas including finance, neuroscience, sociology, and others. The linear Granger [1] approach has become the most popular method for causality analysis in time series due to its computational simplicity. A variable X is said to Granger-cause another variable Y if the prediction error of Y in a linear regression model including its own past values and the past values of X as predictors is less (in some suitable sense) than the prediction error of Y in a linear regression model including only its own past values. Testing for Granger causality is commonly provided by the F-test or Wald test. X Granger-causes Y means that the variable X was found to be helpful for forecasting variable Y. So, the causality in the Granger sense identifies with predictability and precedence. Note, that the procedure for analyzing the causal link between time series can be theoretically used also for time series actually not generated by a linear regression model. The most important issue for using the Granger causality analysis is a validation of the model assumptions, at least

that time series are stationary and no unit roots are presented in the model. If a model assumption is not satisfied, the result of the test for Granger causality is not approved.

Extensions of Granger-causality to non-linear cases have been explored e.g. in [2], [3] and [4]. Also, the causality detection methods that do not directly follow the traditional Granger causality methodology have been proposed, e.g. [5], [6], [7]. However, it is not clear how to validate the goodness of using a method for a particular data and none of the published causality detection methods is appropriate for analyzing causal influence between time series in general, see e.g. [8]. Furthermore, extensive computations and no straightforward interpretation of their numeric results decrease the usefulness of some of these procedures. It can be concluded that identifying a causal relationship among simultaneously acquired signals is still not a satisfactorily closed issue, even for the bivariate case.

Here presented modified Granger causality in selected neighborhoods (MGC) is based on a local approximation of the reconstructed dynamics by a linear regression model. A similar approach to ours has appeared in [2], [9]. The presented procedure differs from the mentioned methods in the way how the local neighborhoods are selected and the most important difference is in using statistical quantity for exploring the causal inference between time series. In our procedure, the direct causal influence of X on Y is evaluated by determining the proportion of variation of the modeled variable, fitted by the linear regression model including past values of both variables, explained by the past of the second variable only. These proportions are calculated only on those local neighborhoods, where the linear regression model fits the observations. The goodness of fit in neighborhoods is assessed by the coefficient of determination, R-squared (R^2). Finally, MGC is obtained as the average of these proportions from suitable neighborhoods over the attractor.

We examined suggested measure on numerically generated nonlinear time series with known patterns of interaction, originally used in the comparative study of causality detection methods [8]. Six test data sets were analyzed: a unidirectional connection of two Hénon systems, a unidirectional connection of chaotic systems of Rössler and Lorenz type and of two different Rössler systems, an example of bidirectionally connected two-species systems, and a fishery model as an example of two correlated observables without a causal relationship.

2 Methodology

Consider two systems X and Y, represented by the time series x and y, respectively. The classical Granger causality starts with fitting the series by a bivariate autoregressive model ($VAR(p)$), where the Akaike information criterion (AIC) or the Schwartz-Bayesian information criterion (BIC) is usually used to determine the order p. In this study, the predictors, which are included in the modeling, are chosen by employing Takens' time-delay embedding [10]. The reconstructed state spaces $\widetilde{X}$ and $\widetilde{Y}$ corresponding to the time series are formed by the vectors

defined as

$$x^{m_x,\tau_x}(t) = (x(t), x(t-\tau_x), \ldots, x(t-(m_x-1)\tau_x))^T, \tag{1}$$

$$y^{m_y,\tau_y}(t) = (y(t), y(t-\tau_y), \ldots, y(t-(m_y-1)\tau_x))^T, \tag{2}$$

where m_x, m_y are the embedding dimensions and τ_x, τ_y are the time delays. The most common practice to determine the reconstruction parameters is to take the delay as the first minimum of the mutual information between the delayed components [11] and the embedding dimension is estimated by the false near neighbor technique [12]. However, in this study, knowing the complexity of the generating systems, we used the lowest possible dimensions for the reconstruction. The joint dynamics of the time series can be expressed by points

$$z^{m_x,\tau_x,m_y,\tau_y}(t) = (x^{m_x,\tau_x}(t)^T, y^{m_y,\tau_y}(t)^T)^T, \tag{3}$$

in the $(m_x + m_y)$-dimensional space $\widetilde{Z}$. The full linear regression models of current values of the time series including the past of both time-delay embedded vectors are defined as

$$\begin{aligned} x(t) &= c_{xy} + \sum_{j=1}^{m_x} a_{xx} x(t-j\tau_x) + \sum_{j=1}^{m_y} a_{xy} y(t-j\tau_y) + \epsilon_{xy}(t), \\ y(t) &= c_{yx} + \sum_{j=1}^{m_x} a_{yx} x(t-j\tau_x) + \sum_{j=1}^{m_y} a_{yy} y(t-j\tau_y) + \epsilon_{yx}(t), \end{aligned} \tag{4}$$

where c_{xy}, c_{yx} are intercepts, $\epsilon_{xy}(t)$, $\epsilon_{yx}(t)$ are the prediction error terms and their magnitudes can be evaluated by their variances, i.e. $var(\epsilon_{xy})$, $var(\epsilon_{yx})$. The autoregressive models of current values of the time series with own past time-delay vectors are of the following form

$$\begin{aligned} x(t) &= c_{xx} + \sum_{j=1}^{m_x} a_x x(t-j\tau_x) + \epsilon_{xx}(t), \\ y(t) &= c_{yy} + \sum_{j=1}^{m_y} a_y y(t-j\tau_y) + \epsilon_{yy}(t), \end{aligned} \tag{5}$$

where c_{xx}, c_{yy} are intercepts, $\epsilon_{xx}(t)$, $\epsilon_{yy}(t)$ are the prediction error terms and their variances are denoted as $var(\epsilon_{xx})$, $var(\epsilon_{yy})$. The intercept-only models, which have no predictors variable, of current values of the time series are of the following form

$$\begin{aligned} x(t) &= \bar{x} + \epsilon_x(t), \\ y(t) &= \bar{y} + \epsilon_y(t), \end{aligned} \tag{6}$$

where $\epsilon_x(t)$, $\epsilon_y(t)$ are the prediction error terms, $\bar{x}$ and $\bar{y}$ is mean of x and y, respectively. The MGC is based on fitting linear regression models (4, 5, 6) in the selected neighborhoods of the reconstructed dynamics.

2.1 Local neighborhoods

For identifying a causal influence of X on Y, a fit of modeled variable Y by full linear regression model in a neighborhood is firstly checked in our procedure. For a point in the (m_x+m_y)-dimensional subspace $\widetilde{Z}$, a corresponding neighborhood is created by its k-nearest points in $\widetilde{Z}$ and R-squared (R^2) is used to evaluate the model quality. R^2 close to one indicates a good model fit, i.e. a large amount of response variability is explained by the linear regression model.

Let $z^{m_x,\tau_x,m_y,\tau_y}(t_{[0]})$ be a point in $\widetilde{Z}$ and $z^{m_x,\tau_x,m_y,\tau_y}(t_{[1]})$, $z^{m_x,\tau_x,m_y,\tau_y}(t_{[2]})$, $\ldots$, $z^{m_x,\tau_x,m_y,\tau_y}(t_{[k]})$ be its k-nearest neighbors. Define the sum of squares

$$\begin{aligned} s^2_{xy} &= \sum_{j=0}^{k} \hat{\epsilon}^2_{xy}(t_{[j]}),\ s^2_{xx} = \sum_{j=0}^{k} \hat{\epsilon}^2_{xx}(t_{[j]})\ s^2_x = \sum_{j=0}^{k} \hat{\epsilon}^2_x(t_{[j]}), \\ s^2_{yx} &= \sum_{j=0}^{k} \hat{\epsilon}^2_{yx}(t_{[j]}),\ s^2_{yy} = \sum_{j=0}^{k} \hat{\epsilon}^2_{yy}(t_{[j]}), s^2_y = \sum_{j=0}^{k} \hat{\epsilon}^2_y(t_{[j]}), \end{aligned} \tag{7}$$

where $\hat{\epsilon}_{xy}$, $\hat{\epsilon}_{yx}$, $\hat{\epsilon}_{xx}$, $\hat{\epsilon}_{yy}$, $\hat{\epsilon}_x$, $\hat{\epsilon}_y$ are estimates of ϵ_{xy}, ϵ_{yx}, ϵ_{xx}, ϵ_{yy}, ϵ_x, ϵ_y defined in (4, 5, 6), respectively. Then, the causal link X to Y is explored in the neighborhood corresponding to $z^{m_x,\tau_x,m_y,\tau_y}(t_{[0]})$ if the coefficient of determination, defined as

$$R^2_{y/\{x,y\}} = 1 - s^2_{yx}/s^2_y, \tag{8}$$

is satisfactorily close to one. A value $R^2_{y/\{x,y\}}$ is a ratio of the explained sum of squares by a fitted full linear regression model containing past of both variables X, Y as predictors and the total sum of squares of the modeled variable Y. Analogously, the causal link Y to X is explored in the neighborhood corresponding to $z^{m_x,\tau_x,m_y,\tau_y}(t_{[0]})$ if the coefficient of determination $R^2_{x/\{x,y\}} = 1 - s^2_{xy}/s^2_x$ is satisfactorily close to one. A value $R^2_{x/\{x,y\}}$ is a ratio of the explained sum of squares by a fitted full linear regression model containing past of both variables X, Y as predictors and the total sum of squares of the modeled variable X.

2.2 Modified Granger causality index

The idea behind the Granger causality is well comprehended. X Granger-causes Y, if the autoregressive prediction error of y is reduced by inclusion of past values of x, that is, $var(\epsilon_{yy}) > var(\epsilon_{yx})$ in a suitable sense. The magnitude of the classical Granger causality X to Y is defined to be log-likelihood ratio $F_{X\to Y} = \ln(var(\epsilon_{yy})/var(\epsilon_{yx}))$. It is important to note that it is meaningless to compare the magnitudes from another couple of series since $F_{X\to Y}$ is not scaled on a range. In the extended Granger causality [2] the causal influence of X on Y is expressed by the variable

$$\delta_{X\to Y} = 1 - var(\epsilon_{yx})/var(\epsilon_{yy}), \tag{9}$$

which is more pragmatic quantity from this point of view. The value of $\delta_{X\to Y}$ indicates a proportion of the variance of y in the autoregressive model, which is

not explained by past of y itself and can be explained by added past of x as predictors to a linear representation of the observed processes. Granger causality is focused on improving the prediction error of modeled variable in the autoregressive model, but the value of the prediction error alone is out of interest. It means, that if $(var(\epsilon_{yy}), var(\epsilon_{yx})) = (10, 1)$ or $(var(\epsilon_{yy}), var(\epsilon_{yx})) = (10^{-4}, 10^{-5})$, then $\delta_{X\to Y} = 0.9$ in both cases, e.g. 90% of unexplained variation of Y by its own past is expressed by implementing the past of X to linear representation of Y and the causal link X to Y is indicated. Consequently, $\delta_{X\to Y}$ can lead to the incorrect indication of a causal effect between time series. In fact, the recent comparative study [8] of various approaches for detection of causality on different data sets demonstrated that the extended Granger causality and the Kernel Granger causality often detect false causality.

The suggested novel estimator is based on specifying the proportion of the variation of the modeled variable explained actually by past values of the second variable in full linear regression model. Furthermore, information about the quality of fit of the full linear regression model is contained in the novel estimator. In particular, the causal influence of X on Y is evaluated by the variable

$$\lambda_{X\to Y} = \frac{\lambda'_{X\to Y}}{R^2_{y/\{x,y\}}} = \frac{s^2_{yy} - s^2_{yx}}{s^2_y - s^2_{yx}}, \text{ where } \lambda'_{X\to Y} = \frac{s^2_{yy} - s^2_{yx}}{s^2_y} \tag{10}$$

in a suitable neighborhood. A value $\lambda'_{X\to Y}$ is a ratio of the explained sum of squares of y from the autoregressive model by the full linear regression model and the total sum of squares of variable y. Thus, the value $\lambda_{X\to Y}$ indicates the proportion of the variation of the response variable y, explained by the linear regression model including past of x and y as predictors, actually predicted by x values only. The magnitude of the suggested modified Granger causality X to Y, denoted $\Lambda_{X\to Y}$, is the average of the proportions $\lambda_{X\to Y}$ from all suitable neighborhoods on the attractor. Here, as the significant presence of the causal influence of X on Y, a value of the MGC index $\Lambda_{X\to Y}$ greater than 0.01 is considered. $\Lambda_{X\to Y} \geq 0.01$ means, that more than 1% of the explained variation of y by the full linear regression model is actually expressed by past of x in the linear representation. In a similar way, the magnitude of the MGC of Y on X denoted $\Lambda_{Y\to X}$ can be defined.

3 Numerical Experiments

Six data sets, each of 20 000 points, with known causal links were analyzed: a unidirectional connection of two Hénon systems, a unidirectional connection of chaotic systems of Rössler and Lorenz type and of two different Rössler systems, an example of bidirectionally connected two-species systems, a fishery model as an example of two correlated observables without a causal relationship, and an example of mediated causality. For more details about generating the data see [8].

Time delay for the studied examples was set to $\tau_x = \tau_y = \tau$, $\tau = 1$. Neighborhoods in the reconstructed dynamics consisted of at least 30 points. R^2 had

to be at least 0.95 and the number of points in a suitable neighborhood was increased as long as R^2 did not decrease.

In all experiments, we compared the novel measure with the extended Granger causality (EGC), which is also based on a local approximation of time delay embedding reconstruction of state space by a linear regression model. The local neighborhood for a point in $\widetilde{Z}$ is created by points of $\widetilde{Z}$ within specified distance r. Then, the EGC index for the causal link X to Y is the average of calculated values of the variable (9) in such local neighborhoods. The distance r was chosen to obtain comparatively large neighborhoods to those created in the MGC.

3.1 Hénon $\longrightarrow$ Hénon

In the first example, we studied two unidirectionally coupled identical Hénon maps, see e.g. [13], [14], [15]. The first two lines correspond to the driving system X, and the last two equations describe the response system Y:

$$\begin{aligned} x_1(t+1) &= 1.4 - x_1^2(t) + 0.3x_2(t) \\ x_2(t+1) &= x_1(t) \\ y_1(t+1) &= 1.4 - [Cx_1(t)y_1(t) + (1-C)y_1^2(t)] + 0.3y_2(t) \\ y_2(t+1) &= y_1(t), \end{aligned} \tag{11}$$

where the strength of the coupling is controlled by the parameter C. The systems are coupled through a one-way driving relationship between variables x_1 and y_1. It follows that causal influence of x_1 on y_1 should increase if the strength of coupling C increases. We varied C from 0.08 to 0.72 with increments of 0.08. From other studies [16], it is known that x_1 and y_1 will enter identically synchronized state for $C \leq 0.7$.

The embedding dimension for both series was set up to $d_{x_1} = d_{y_1} = 2$ and the distance $r = 0.2$. Fig. 1 shows results of our experiments. We observe that the MGC successfully indicates the absence of a causal influence of y_1 on x_1 at any C, but the EGC incorrectly indicates the bidirectional causal link between investigated variables.

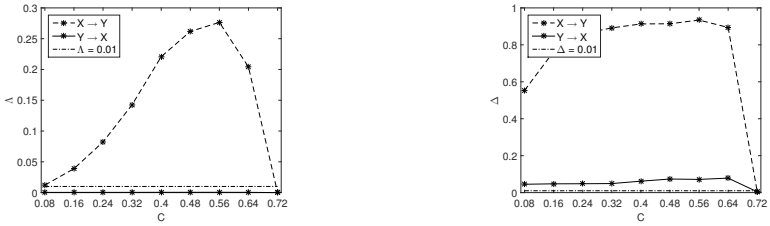


Fig. 1. Unidirectionally coupled Hénon systems [Eqs. (11)]: $X := x_1$, $Y := y_1$. The modified Granger causality index Λ (on the left) and the extended Granger causality index Δ (on the right).

3.2 Rössler ⟶ Lorenz

The next example is a unidirectional coupling of two different chaotic systems, see e.g. [17], [18]. The Rössler system X drives the Lorenz system Y:

$$\begin{aligned}
\dot{x}_1 &= -6(x_2 + x_3) \\
\dot{x}_2 &= 6(x_1 + 0.2x_2) \\
\dot{x}_3 &= 6[0.2 + x_3(x_1 - 5.7)] \\
\dot{y}_1 &= 10(-y_1 + y_2) \\
\dot{y}_2 &= 28y_1 - y2 - y_1y_3 + Cx_2 \\
\dot{y}_3 &= y_1y_2 - \frac{8}{3}y_3.
\end{aligned} \tag{12}$$

The coupling strength is controlled by the parameter C, which was varied from 0 to 4 with a step of 0.4. Generalized synchronization seems to take place just before coupling $C = 3$ [19]. For testing purposes, a causal link between variable x_2 of the driving Rössler system and variable y_2 of the driven Lorenz system was investigated.

The embedding dimension for both series was set up to $d_{x_2} = d_{y_2} = 3$ and the distance $r = 3$. Fig. 2 shows results of our experiments. We observe that the MGC successfully indicates the absence of a causal influence of y_2 on x_2 at any C. As in the previous example, the EGC incorrectly indicates the bidirectional causal link between the investigated variables.

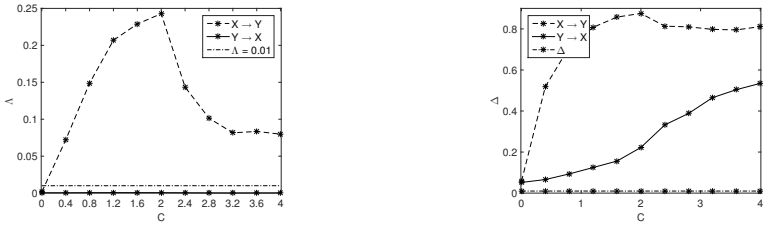


Fig. 2. Unidirectionally coupled Rössler-Lorenz system [Eqs. (12)]: $X := x_2$, $Y := y_2$. The modified Granger causality index Λ (on the left) and the extended Granger causality index Δ (on the right).

3.3 Rössler 0.5 ⟶ Rössler 2.5

In the next testing example, unidirectionally coupled nonidentical Rössler systems with coupling strengths C [15] were used:

$$\begin{aligned}
\dot{x}_1 &= -\omega_1 x_2 - x_3 \\
\dot{x}_2 &= \omega_1 x_1 + 0.15 x_2 \\
\dot{x}_3 &= 0.2 + x_3(x_1 - 10) \\
\dot{y}_1 &= -\omega_2 y_2 - y_3 + C(x_1 - y_1) \\
\dot{y}_2 &= \omega y_1 + 0.72 y_2 \\
\dot{y}_3 &= 0.2 + y_3(y_1 - 10),
\end{aligned} \tag{13}$$

where $\omega_1 = 0.5$, $\omega_2 = 2.515$. The causal link between variables x_1 and y_1 was investigated. The coupling strength C was chosen from 0 to 1 with a step size of 0.2. The synchronization takes place at a coupling of about 1. In addition, these systems also differ in that the first is chaotic, whereas the other is quasiperiodic [15].

The embedding dimension for both series was set up to $d_{x_1} = d_{y_1} = 3$ and the distance $r = 3$. Fig. 3 shows results of our experiments. We observe that the MGC incorrectly indicates no causal influence of x_1 on y_1 and the EGC incorrectly indicates the causal link y_1 to x_1.

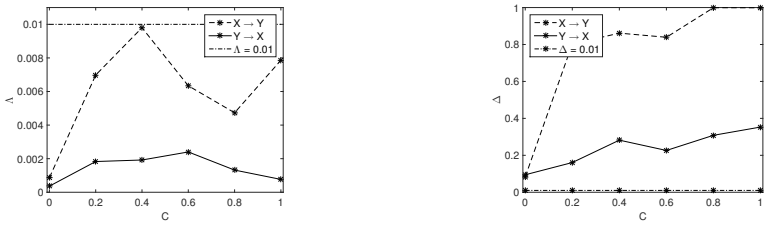


Fig. 3. Unidirectionally coupled Rössler systems [Eqs. (13)]: $X := x_1$, $Y := y_1$. The modified Granger causality index Λ (on the left) and the extended Granger causality index Δ (on the right).

3.4 Bidirectional Two-Species Model

The next example represents a bidirectional causality between two coupled logistic difference equations [19]:

$$\begin{aligned}
x(t+1) &= x(t)[3.78 - 3.78x(t) - 0.07y(t)] \\
y(t+1) &= y(t)[3.77 - 3.77y(t) - 0.08x(t)].
\end{aligned} \tag{14}$$

The embedding dimension for both series was set up to $d_x = d_y = 2$ and the distance $r = 0.02$. The results of our experiments are presented in Table 1. We observe that both measures correctly indicate bidirectional causality.

3.5 Fishery Model

The fishery model, represents the situation of a correlation between systems that can be falsely declared as causality. We studied a standard logistic model of two noninteracting fish populations sharing common environment (e.g., weather) [6], [7]:

$$\begin{aligned} R_x(t+1) &= x(t)\{3.1[1-x(t)]\}exp(0.5Z(t)) \\ R_y(t+1) &= y(t)\{2.9[1-y(t)]\}exp(0.6Z(t)) \\ x(t+1) &= 0.4x(t) + max[R_x(t-3), 0] \\ y(t+1) &= 0.35y(t) + max[R_y(t-3), 0]. \end{aligned} \tag{15}$$

The variables x and y denote the sizes of the fish populations, R_x and R_y are the recruitments of the populations, and Z is an environmental influence represented here by a red noise defined as:

$$Z(i) = p \sum_{j=i-14}^{i} Z^{'}(j),\ Z^{'} \sim N(0,1),$$

where the variable p was chosen so that $var(Z)$ was equal to 1.

The embedding dimension for both series was set up to $d_x = d_y = 5$ and the distance $r = 0.6$. The results of our experiments are presented in Table 1. We observe that the MGC successfully indicates the absence of the direct causal link between the investigated variables. The EGC incorrectly indicates the bidirectional causal link between variables.

3.6 Mediated Link

In the last example, two uncoupled subsystems X and Y are driven by the third one, Z. Only the case with coupling of $C = 0.06$ was tested for which Z and X were already synchronized, and this was reflected as a mediated causal effect from X to Y. The equations are as follows:

$$\begin{aligned} \dot{z}_1(t) &= -\omega_1 z_2(t) - z_3(t) \\ \dot{z}_2(t) &= \omega_1 z_1(t) + 0.15 z_2(t) \\ \dot{z}_3(t) &= 0.2 + z_3(t)(z_1(t) - 10) \\ \dot{x}_1(t) &= -\omega_2 x_2(t) - x_3(t) + C[z_1(t) - x_1(t)] \\ \dot{x}_2(t) &= \omega_2 x_1(t) + 0.15 x_2(t) \\ \dot{x}_3(t) &= 0.2 + x_3(t)[x_1(t) - 10] \\ \dot{y}_1(t) &= -\omega_3 y_2(t) - y_3(t) + C[z_1(t) - y_1(t)] \\ \dot{y}_2(t) &= \omega_3 y_1(t) + 0.15 y_2(t) \\ \dot{y}_3(t) &= 0.2 + y_3(t)[y_1(t) - 10]. \end{aligned} \tag{16}$$

The frequencies ω_1, ω_2, and ω_3 were selected to have the following, approximately unit, ratios: $\omega_1 = 0.985$, $\omega_2 = 1$, and $\omega_3 = 1.015$, respectively. For testing, a causal link between x_1 and y_1 was investigated.

The embedding dimension for both series was set up to $d_{x_1} = d_{y_1} = 3$ and the distance $r = 2$. The results of our experiments are presented in Table 1. As in the previous example, we observe that the MGC correctly indicates the absence of the direct causal link between variables and that the EGC incorrectly indicates the bidirectional causal effect.

Table 1. The modified Granger causality index (Λ) and the extended Granger causality index (Δ) for the bidirectional two-species data, the fishery data, and the mediated link data.

Data	$\Lambda_{X\to Y}$	$\Lambda_{Y\to X}$	$\Delta_{X\to Y}$	$\Delta_{Y\to X}$
Bidir	0.016811	0.017117	0.99792	0.99826
Fishery	0.0062587	0.0019754	0.061718	0.10703
Mediated Link	0.00071925	0.00086107	0.1591	0.14025

4 Conclusions

We present a novel causality estimator for bivariate time series, the modified Granger causality in selected neighborhoods (MGC). It is based on fitting linear regression models in local neighborhoods of the reconstructed joint dynamics of the time series and reflects the proportion of variation of the modeled variable, fitted by the linear regression model including past values of both variables, actually explained by the past of the second variable in the linear representation. The effectiveness of the suggested causal measure was studied on numerically generated nonlinear time series. Performance of the Granger vector autoregressive test, the extended Granger test, the kernel version of the Granger test, the conditional mutual information (transfer entropy), the evaluation of cross-mappings between state spaces, and an assessment of predictability improvement due to the use of mixed predictions was analyzed on the similar data sets in a recently published extensive comparison study [8]. The results showed that the MGC analysis is superior or comparable to these alternative measures in quantifying the asymmetric causal interdependent bivariate time series data. Except of the determination of the reconstruction parameters, the evaluation of the MGC requires only least squares calculation. The computational properties and the experimental results indicate that the MGC is potentially first choice technique in many cases to reveal symmetric or asymmetric dependences between coupled bivariate time series.

5 Acknowledgments.

The work was supported by the Slovak Research and Development Agency, project APVV-15-0295, and by the Scientific Grant Agency VEGA of the Min-

istry of Education of the Slovak Republic and the Slovak Academy of Sciences, by the projects VEGA 2/0011/16 and VEGA 2/0054/18.

References

1. Granger, C. W.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods, Econometrica 37, 424–438 (1969).
2. Chen, Y., Rangarajan, G., Feng, J., Ding, M.: Analyzing Multiple Nonlinear Time Series with Extended Granger Causality. Phys. Lett. A 324, 26–35 (2004)
3. Marinazzo, D., Pellicoro, M., Stramaglia, S.: Kernel Method for Nonlinear Granger Causality. Phys. Rev. Lett. 100:144103 (2008)
4. Nicolaou, N., Constandinou, T.G.: A Nonlinear Causality Estimator Based on Non-Parametric Multiplicative Regression. Frontiers in Neuroinformatics 10:19 (2016)
5. Shannon, C. E.: A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379–423 (1948).
6. Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., Munch, S.: Detecting Causality in Complex Ecosystems. Science 338, 496–500 (2012).
7. Krakovská, A., Hanzely, F.: Testing for Causality in Reconstructed State Spaces by an Optimized Mixed Prediction Method. Phys. Rev. E 94:052203 (2016).
8. Krakovská, A., Jakubík, J., Chvosteková, M. et al.: Comparison of Six Methods for the Detection of Causality in a Bivariate Time Series. Phys. Rev. E 97:042207 (2018)
9. Freiwald, W.A., Valdes, P., Bosch, J. et al.: Testing Non-linearity and Directedness of Interactions Between Neural Groups in the Macaque Inferotemporal Cortex. J. of Neurosci. Methods 94, 105–119 (1999)
10. Takens, F.: Detecting Strange Attractors in Turbulence. In: Rand, D.A. and Young, L.-S. (eds.) Dynamical Systems and Turbulence, Lecture Notes in Mathematics, vol. 898, pp. 366-381. Springer, Verlag (1981)
11. Fraser, A. M., Swinney, H. L.: Independent Coordinates for Strange Attractors from Mutual Information. Phys Rev A 33, 1134–1140 (1986)
12. Kennel, M. B., Brown, R., Abarbanel, H. D. I.: Determining Embedding Dimension for Phase-Space Reconstruction Using a Geometrical Construction. Phys Rev A 45, 3403–3411 (1992)
13. Hénon, M. A Two-Dimensional Mapping with a Strange Attractor. Comm. Math. Phys. 50, 69–77 (1976).
14. Schiff, S. J., So, P., Chang, T., Burke, R. E., Sauer, T.: Detecting Dynamical Interdependence and Generalized Synchrony through Mutual Prediction in a Neural Ensemble. Phys. Rev. E 54, 6708 (1996).
15. Paluš, M., Vejmelka, M.: Directionality of Coupling from Bivariate Time Series: How to Avoid False causalities and missed connections. Phys. Rev. E 75, 056211 (2007).
16. Schreiber, T.: Measuring Information Transfer, Phys. Rev. Lett. 85, 461-464 (2000)
17. Paluš, M., Komárek, V., Hrnčíř, Z., Štěrbová, K.: Synchronization as Adjustment of Information Rates: Detection from Bivariate Time Series. Phys. Rev. E 63:046211 (2001).
18. Krakovská, A., Mezeiová, K., Budáčová, H.: Use of False Nearest Neighbours for Selecting Variables and Embedding Parameters for State Space Reconstruction. J. Complex Syst. 2015, 932750 (2015).
19. Ye, H., Deyle, E. R., Gilarranz, L. J., Sugihara, G.: Distinguishing Time-Delayed Causal Interactions Using Convergent Cross Mapping. Sci. Rep. 5:14750 (2015).

State of Charge Depended Modeling of an Equivalent Circuit of Zinc Air Batteries Using Electrochemical Impedance Spectroscopy

Andre Loechte, Ole Gebert, Ludwig Horsthemke, Daniel Heming, and Peter Gloesekoetter

University of Applied Sciences Muenster,
Dept. of Electrical Engineering and CS.,
Stegerwaldstr. 39, 48565 Steinfurt, Germany
{a.loechte@fh-muenster.de}
www.fh-muenster.de

Abstract. Metal air batteries provide a high energy density as the cathodic reaction uses the surrounding air. Different metals can be used but zinc is very promising due to its disposability and nontoxic behavior. State estimation is quite complicated as the voltage characteristic of the battery is rather flat. Especially estimating the state of charge is important as a secondary electrolysis process during overcharging can lead to an unsafe state. Another technique for state estimation is the electrochemical impedance spectroscopy. Therefore, this paper describes the process of setup and measuring a time series of impedance spectra at known states of charge. Then these spectra are used to derive an equivalent circuit. Finally the development of the circuit's parameter are analyzed to extract most important parameters.

Keywords: zinc air battery, electrochemical impedance spectroscopy, state of charge, equivalent circuit

1 Introduction and Problem

Zinc air batteries provide a high energy density as the cathodic reaction uses the surrounding air. Typical characteristics of zinc air batteries are flat charging and discharging curves. Compared to a lithium-ion battery, the voltage difference between a fully charged and an empty battery is 10 times smaller[1]. On one hand this is an advantage for the following power electronics which can be optimized for smaller voltage ranges. On the other hand, state estimation of the system becomes more complex as the voltage level is not sufficient to identify the state of charge. Another possibility is to calculate the state of charge by measuring and integrating the battery's current. The unknown self discharge and aging behavior lead to huge uncertainties, though.
Moreover, there is no end-of-charge voltage which indicates a full battery. While outer voltage and impressed current stay the same, hydrogen and oxygen are

produced by a smoothly starting electrolysis instead of deoxidizing the zinc oxide[2]. So the most important factor is to identify if the battery is full as the hydrogen oxygen mixture can get potentially explosive after some time of overcharging. In this context, the electrochemical impedance spectroscopy tends to be a promising candidate. The procedure of impedance spectroscopy in a galvanostatic way is to apply a sinus current to the battery while measuring the resulting voltage. Both signals are then used to calculate the ac impedance at a certain frequency. This procedure is repeated several times for different frequencies resulting in an impedance spectrum. Among other parameters, these spectra depend on the state of charge and can therefore be used in a battery management system[3][4][5].

2 Procedure

In order to verify this concept, a zinc air battery with a nominal capacity of 100 A h is charged and discharged cyclically while impedance spectra are measured each 30 min. Furthermore the current state of charge is determined and logged by integrating the applied charging and discharging current, respectively. The self discharge rate is negligible as the battery is used permanently. This results in a set of impedance spectra at known states of charge. These spectra are then used to determine parameters of an equivalent circuit of the battery.

2.1 Test setup

The setup that is used to measure the impedance spectra is shown in Figure 1. A PC oscilloscope with an arbitrary waveform generator (AWG) and two voltage measurement units(In 1 and In 2) is controlled by a computer using MATLAB. A measurement starts with a sine wave which is generated by the waveform generator. This voltage signal is the input for a current controller that generates the sinus current signal that is applied to the battery under test (BUT). The negative feedback of the current controller is created by a shunt resistor. Its voltage is also measured by one input of the oscilloscope so that we have information about the actual current of the battery cell. The resulting voltage response of the battery is measured as well. Since the oscilloscope got a programmable gain amplifier[6], an offset unit is used to increase the precision of the measurement. By eliminating the dc offset of the battery, the smallest measuring range of the oscilloscope can be used. The ac component could barely measured without this compensation as its amplitude is quite small (about ± 10 mV) compared to the dc offset which is up to 2.45 V for charging zinc air cell. This compensation results in an effective number of 5.67 bits to measure the ac voltage response, as the minimal voltage range of the 8 bit voltage measuring unit is ± 50 mV. This setup is based on a development named picoEIS of the institute for Power Electronics and Electrical Drives of the RWTH Aachen university[7]. However, its schematic uses a potentiometer to eliminate the dc offset, so manual controlling is necessary for each measurement. In order to automate the process for time

series measurements, a micro controller with an analog digital converter (ADC) is used. The output voltage of the ADC works as offset and is determined by a successive approximation.

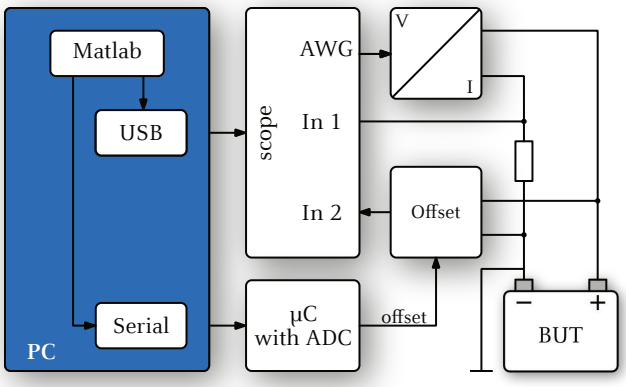


Fig. 1. Test setup that is used for measuring the impedance spectra of the zinc air battery. The dc current offset (not shown) is applied in a parallel connection to the battery under test.

Figure 2 shows the resulting process for measuring the time series of impedance spectra in a flowchart. There are two separate loops. The left one is used to cyclically charge and discharge the battery by controlling a power supply in constant current mode or a battery discharging unit. Over the entire term, the battery's voltage and current are logged. The state of charge is approximated by integrating the measured current and is logged as well. These parameters are also used to determine when changes from charging to discharging and vice versa are necessary. The right loop measures the impedance spectrum each 30 min for a fixed set of frequencies. Each measurement starts with eliminating the battery's dc voltage offset. Then the ac current signal is applied and resulting current and voltage signals of the battery are measured for 3 periods or 1 s, whatever is longer. A discrete Fourier transform is used to calculate the impedance.Then, the determined impedance is logged together with the raw measurement data and the battery's voltage, current and state of charge.

2.2 Model

The resulting spectra are used to derive an electrical battery model whose parameters depend on the state of charge. Separate models for charging and discharging are used, as the utilized battery has three electrodes. While the zinc electrode is

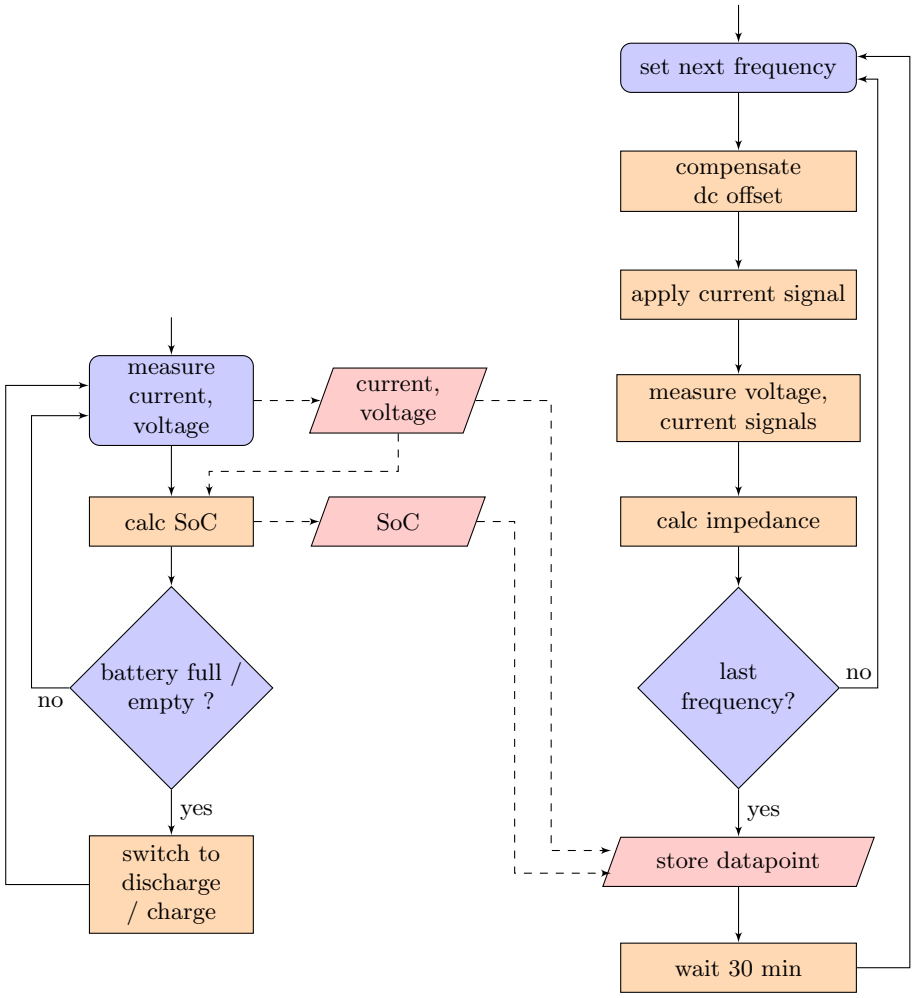


Fig. 2. Flowchart that describes the process of training and validation of one combination of number of frequency and neurons

used for both processes, there are separate electrodes for charging and discharging. Like in primary bottom cells, a gas diffusion electrode made of graphite is used for discharging. During charging process a separate nickel grid is used. Both models consist of three parts, electrolyte, anode and cathode, respectively. The electrolyte works as an ion conductor and is approximated by a resistor R_{el}[8]. Anode and cathode consist of three components. The capacitor C_{dl} describes the double layer capacity of the electrode while the resistor R_{ct} approximates the charge-transfer resistance[8][9]. The diffusion process is approximated by a Nernstian diffusion. Therefore, Ficks second law is solved using the boundary condition that the gradient of the concentration with respect to the location is a complex rotating pointer[10]. This results in an impedance

$$Z_{Nernst} = R_d \frac{[\sinh(x) + \sin(x)] - j[\sinh(x) - \sin(x)]}{x[\cosh(x) + \cos(x)]} \ , \tag{1}$$

where x is

$$x = \sqrt{\frac{2\omega}{K}} \ . \tag{2}$$

The diffusion component describes the behavior at two different states of aggregation. So during charging there is one change between nickel net and electrolyte and one between electrolyte and zinc anode, resulting in one diffusion process per electrode (see Figure 3).

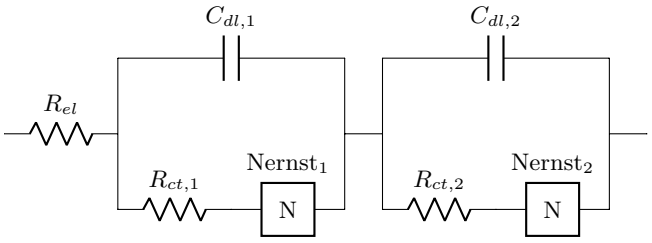


Fig. 3. Equivalent circuit of the zinc air battery during charging process.

During discharging a gas diffusion electrode is used. Since it conjuncts three different states of aggregation, here air, graphite and electrolyte, a second diffusion process is added to the cathodic part of the equivalent circuit (Figure 4)[9].
The parameter of the corresponding equivalent circuit are varied so that the sum of the squared error E_{square}

$$E_{square} = \sum_{i=1}^{n} \left(Z_{measured}(f_i) - Z_{modell}(f_i)\right)^2 \ . \tag{3}$$

is minimized using all measured frequencies f_i of one measured spectrum. For this Matlab is used. As the structure of the equivalent circuit of anode and

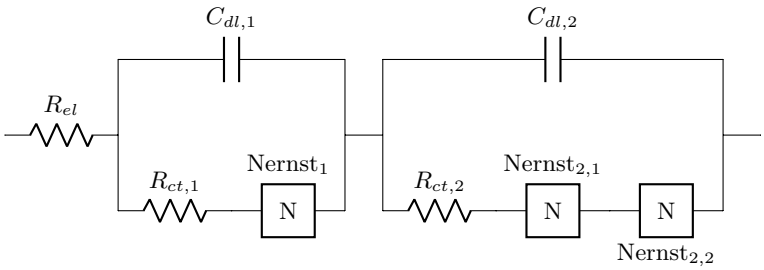


Fig. 4. Equivalent circuit of the zinc air battery during discharging process.

cathode are equal, the solution of the optimization algorithm might vary the counterparts of the circuit's components. So that components with index 1 belong to the anode in one solution while belonging to the cathode in different run. Therefore, the parameter values of the last spectrum are used as starting parameters for the next spectrum in order to fix the parameters to the corresponding part of the circuit.

3 Results

Selected samples of the measured spectra during discharging are shown in Figure 5. The low-budget setup used is sufficient to measure spectra that differ with respect to the state of charge. As one can see, the variance between the spectra is higher at lower frequencies. Reason for this behavior is probably, that the diffusion processes at the gas diffusion cathode are rather slow. Spectra at higher states of charges tend to have higher absolute values for both imaginary and real part. This can be explained by an increasing charge transfer resistance. At higher frequencies the impedance is mostly defined by the conductivity of the electrolyte as the double layer capacitors of anode and cathode work as short cuts. That's why all spectra end in almost the same impedance value.
Two different spectra of a charging cycle are shown in Figure 6. Here only one semicircle per spectrum can be seen. While the spectrum of a half charged battery is reflected by the equivalent circuit, the performance of a fully charged battery does not fit to the equivalent circuit. Probably the electrolysis is dominant in this case, resulting in a completely different behavior. At low frequencies the impedance spectrum tends to drift away with a constant phase, that can be explained with a Warburg element[11]. Once again, the highest variance exist at low frequencies.

4 Analysis

Figure 7 shows some of the determined parameter values of the equivalent circuit during a discharge cycle. The cycle measurements belong to an already wearing

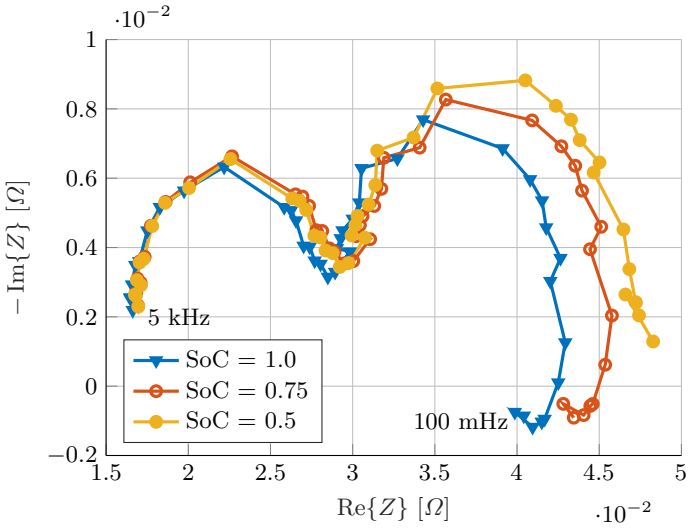


Fig. 5. Some measured sample spectra of a discharging cycle at different state of charges.

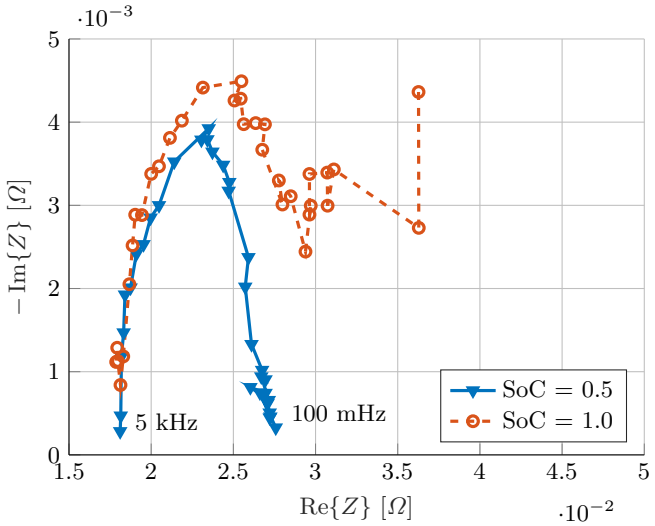


Fig. 6. Two measured spectra of a charging cycle comparing a battery at half charged and at overcharged state.

cell, which explains the end of discharging periods after approx 51 A h. The values are normalized to the mean values in order to clarify its development. Nevertheless the variance is still small and plotted in thousandths. Here $C_{dl,2}$ and $R_{ct,2}$ correspond to the gas diffusion cathode. Its parameter values are rather constant as the oxygen concentration of the surrounding air does not change that much. The absolute change of R_{el} is quite small as the resistivity is the minimum impedance of each spectra. Therefore, the most distinct parameter for evaluating the state of charge is the charge transfer resistance $R_{ct,1}$ of the zinc anode.

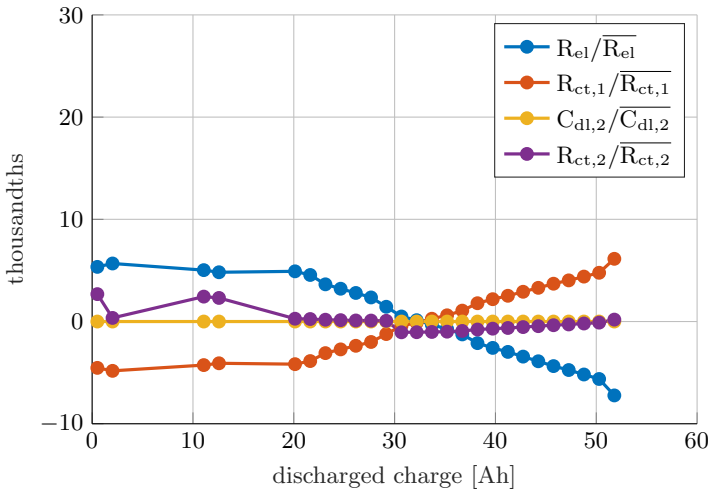


Fig. 7. Development of the normalized parameter of the equivalent circuit during a discharge cycle.

A closer look at $R_{ct,1}$ is taken in Figure 8. The absolute change of the parameter value is rather low, nevertheless a precise trend can be seen. At the beginning of the discharge cycle, the resistance barely changes. The zinc concentration is still high enough to not harm the charge transfer. But it increases linearly after 20 A h have been discharged. A quadratic fit provide a good approximation in both parts of the characteristic.
Figure 9 shows the determined parameter of the model depending on the state of charge during a whole charging cycle. The parameters of the zinc anode indicate a significant change at 68 A h. Keeping in mind that we were able to discharge 51 A h of the battery, this change probably results from the secondary electrolysis process. The electrolysis turns out to be dominant at the end of the charging period. The measurement data of the plotted charging cycle belongs to an already wearing cell. So it is very likely that the battery already lost some of its capacity. The difference between the charged and discharged charge can be explained by the coulombic efficiency. According to that, the electrochemical impedance spectroscopy can definitely be used to determine if a zinc air battery is full. A

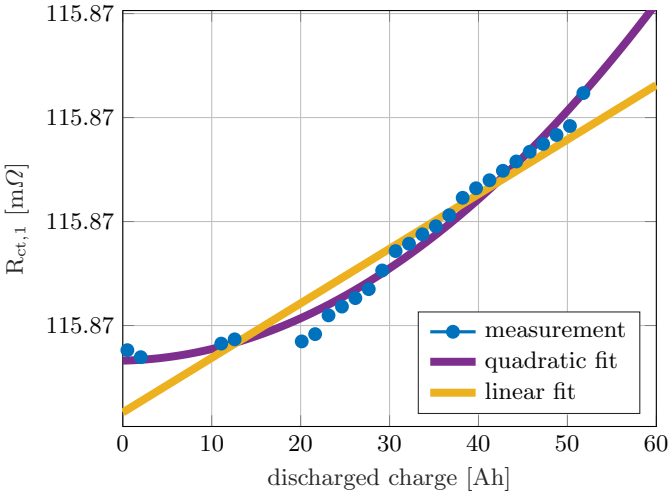


Fig. 8. A closer look to the charge transfer resistance of the anode during a discharge cycle.

closer look to the area where the battery is not yet fully charged does not show distinct dependence of any parameter from the state of charge.
Nevertheless, Figure 10 takes a closer look at $R_{ct,1}$ in the area where the battery is not fully. In this area it shows the most distinct dependence. As one can see, it can be approximated by a linear fit. However, the noise is quiet high, possibly due to the power supply that was used for charging and that depends on a switching regulator. Therefore an improved measurement setup is necessary to explicitly determine the state of charge during charging process. The resistance of $R_{ct,1}$ roughly starts at 24.8 mΩ for an empty battery and decreases to 22.8 mΩ for a full battery. Hence, its absolute value is lower compared to the value during the discharging process.

5 Conclusion and Outlook

The results of the time series showed, that the electrochemical impedance spectra of a zinc air battery definitely differ with respect to the state of charge. An equivalent circuit has been derived from the spectra, revealing that the most distinct parameter for the state of charge is the charge transfer resistance of the anode. While the current setup is sufficient to determine the state of charge while discharging the battery, the noise level during charging is rather high so that only overcharging can be recognized reliably. Therefore, a new setup, shown in Figure 11, is being developed in order to improve the accuracy. Furthermore this setup uses a micro controller resulting in a much smaller embedded system. The accuracy is going to be improved in two ways. On the one hand the creation of the stimulus is revised. The amplitude and offset of the applied current are

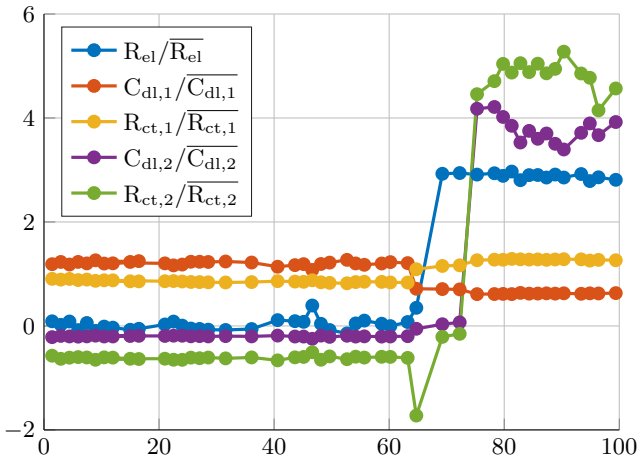


Fig. 9. Development of the normalized parameter of the equivalent circuit during a charging cycle.

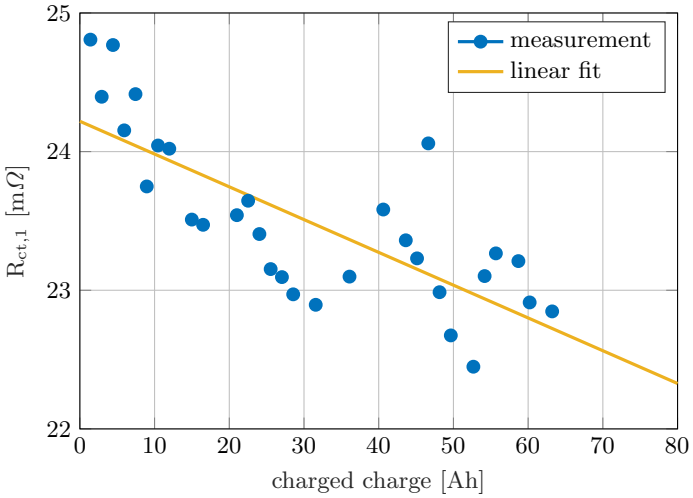


Fig. 10. Development of the charge transfer resistance during a charging cycle while the battery is not overchaged.

controlled separately now, maintaining good accuracy of the sine wave during high dc current offsets. On the other hand, the resolution of the analog digital is increased. Therefore a two channel 24 bit ADC is used. The effective number of bits measuring the ±10 mV ac sine signal is increased to 15 bits, although the dc offset of the battery is not compensated and measured as well. Moreover, the influence of high temperatures and the oxygen content will be examined.

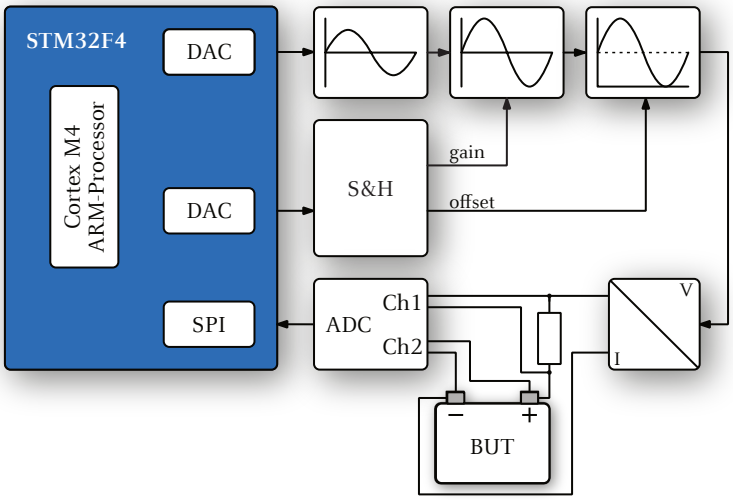


Fig. 11. Revised test setup that will be used for future measurements.

References

1. Expected publication: Loechte, A., Gebert, O. Heming, D., Kallis, K. T., Gloesekoetter, P.: State estimation of zinc air batteries using neural networks. In: Neural Computing and Applications, vol. 521, Springer (2018)
2. Linke, K., Turek, T.: Metall-Luft-Batterien mit Ionischen Flssigkeiten als Elektrolyt, (2013), `https://dechema.de/dechema_media/2940_Schlussbericht-p-4818.pdf`
3. Jimenez Gordon, I. A., Grugeon, S., Takenouti, H., Tribollet, B., Armand, M., Davoisne, C. Debart, A. Laruelle, S.: Electrochemical Impedance Spectroscopy response study of a commercial graphite-based negative electrode for Li-ion batteries

as function of the cell state of charge and ageing. In Electrochimica Acta, vol. 223, pp 63 – 73, (2017)
4. Wang, Q.-K., He, Y.-J., Shen, J.-N., Hu, X., Ma, Z.-F.: State of Charge-Dependent Polynomial Equivalent Circuit Modeling for Electrochemical Impedance Spectroscopy of Lithium-ion Batteries, In: IEEE Transactions on Power Electronics, (2017)
5. Ran, L., Junfeng, W., Haiying, W., Gechen, L.: Prediction of state of charge of Lithium-ion rechargeable battery with electrochemical impedance spectroscopy theory, In: IEEE Conference on Industrial Electronics and Applications, vol. 5, (2010)
6. pico Technology: PicoScope 2200A Series datasheet, https://www.picotech.com/download/datasheets/PicoScope2200ASeriesDataSheet.pdf
7. Kiel, M.: Impedanzspektroskopie an Batterien unter besonderer Beruecksichtigung von Batteriesensoren fr den Feldeinsatz. In: Aachener Beitrge des ISEA, vol. 67, Shaker Verlag (2013)
8. Gamry Instruments: Electrochemical Impedance Spectroscopy: A Primer, (2007), https://www.gamry.com/application-notes/EIS/basics-of-electrochemical-impedance-spectroscopy/
9. Arai, H., Mueller, S., Haas, O.: AC Impedance Analysis of Bifunctional Air Electrodes for Metal-Air Batteries. In: Journal of The Electrochemical Society, vol. 147, (2000)
10. Drossbach, P., Schulz, J.: Elektrochemische Untersuchungen an Kohleelektroden - Die Ueberspannung des Wasserstoffs. In: Electrochemica Acta, vol. 9, (1964)
11. Warburg, E.: Ueber das Verhalten sogenannter arisirbarer Elektroden gegen Wechselstrom. In: Annalen der Physik und Chemie, Neue Folge 67, vol. 303, pp. 493 – 499. (1899)

Cryptanalysis of a Chaos Based Encryption Algorithm for Secure Communication

Salih Ergün

ERGTECH Research Center
Zürich, Switzerland
salih.ergun@ergtech.ch

Abstract. This paper introduces cryptanalysis of a chaos-based encryption algorithm over TCP data packet for secure communication. An attack system is proposed to discover the security weaknesses of the chaos-based encryption algorithm. Convergence of the attack system is proved using master-slave synchronization scheme. Future evaluation of the encryption algorithm is obtained from a scalar time series where the only information available are the structure of the encryption system and a scalar time series observed from the chaotic system. Simulation and numerical results verifying the feasibility of the attack system are given.

1 Introduction

Because of today's emerging technology, the need for computer networks has increased significantly. With the increasing use of the internet, communication has become very easy around the world. After that, the size and rates of data transmitted via computer networks is increasing day by day. Security is one of the important requirements for networking. The information sent and received during data transfer must be protected against external factors that interfere with the system.

Today, individuals and corporations allocate large budgets for data security. Data security is getting worse every day and the losses in data security are steadily increasing. One of the various approaches to ensure data security is to use the dynamic properties of chaotic systems [1, 2]. Chaotic ciphering work attracted more attention than standard ciphering methods. Chaotic systems depend sensitively on initial conditions and parameters [3]. Chaotic systems have become more popular because they can successfully maintain the complexity and propagation of the key components [4] of cryptography such as random number generators (RNGs).

Although the use of discrete-time chaotic maps in the realization of RNG has been widely accepted for a long period of time [5–7], it has been shown during the last decade that continuous-time chaotic oscillators can also be used to realize RNGs [8–10]. In particular, a chaos-based RNG was proposed in [8] and used in an encryption algorithm over TCP data packet for secure communication. In this paper we target the chaos-based encryption algorithm reported in [8]

and further propose an attack system to discover the security weaknesses of the targeted system.

Being aware of any knowledge on the design of the RNG should not provide a useful prediction about the output bit sequence. Even so, fulfilling the requirements for secrecy of cryptographic applications using the RNG dictate three secrecy criteria as a "must": 1. The output bit sequence of the RNG must pass all the statistical tests of randomness; 2. The previous and the next random bit must be unpredictable and; 3. The same output bit sequence of the RNG must not be able to be reproduced [11].

An important principle of modern cryptography is the Kerckhoff's assumption [12], states that the overall security of any cryptographic system entirely depends on the security of the key, and assumes that all the other parameters of the system are publicly known.

Cryptanalysis is the complementary of cryptography. Interaction between these two branches of cryptology form modern cryptography which has become strong only because of security analysis revealing weaknesses in existing cryptographic systems [13].

The strength of a cryptographic system almost depends on the strength of the key used or in other words on the difficulty for an attacker to predict the key. On the contrary to recent RNG design [9, 10], where the effect of noise generated by circuit components was analyzed to address security issue, the target chaos-based encryption system [8] pointed out the deterministic chaos itself as the source of randomness.

The organization of the paper is as follows. In Section 2 the target RNG system is described in detail; In Section 3 an attack system is proposed to cryptanalyze the target system and its convergence is proved; Section 4 illustrates the numerical results with simulations which is followed by concluding remarks.

2 Target System

Chaotic systems are categorized into two groups: discrete-time or continuous-time, respectively regarding on the evolution of the dynamical systems. In target chaos-based encryption system [8], a simple continuous-time chaotic system is utilized as the core of the RNG. This is a three-dimensional chaotic system with no real equilibrium points which is also called hidden attractor chaotic system. The system consists of three different equations as shown in Eqn. 1 with three state variables:

$$\begin{aligned} \dot{x_1} &= ay_1 - x_1 + z_1 y_1 \\ \dot{y_1} &= -bx_1 z_1 - cx_1 + y_1 z_1 + d \\ \dot{z_1} &= e - f{x_1}{y_1} - {x_1}^2 \end{aligned} \tag{1}$$

The equations in 1 generate chaos for different sets of parameters. The chaotic attractor given in Fig.1 is obtained from the numerical analysis of the system with $a = 2.8$, $b = 0.2$, $c = 1.4$, $d = 1$, $e = 10$ and $f = 2$. Target random number generation mechanism is described in [8] by a flow chart. As depicted in [8]

the core chaotic system given in Eqn. 1 was solved by a 4^{th}-order Runge-Kutta method and time series were found.

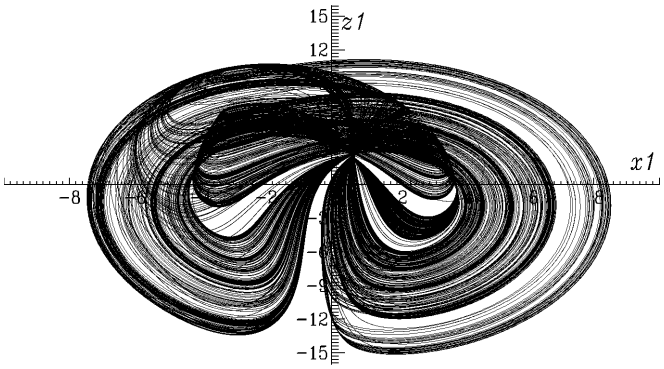


Fig. 1. Numerical analysis results of the chaotic system for $a = 2.8$, $b = 0.2$, $c = 1.4$, $d = 1$, $e = 10$ and $f = 2$.

Then, floating point numbers to be used in encryption are selected at specific intervals in this time series. In the next step, the selected floating-point numbers are converted to a binary number system. When each of floating-point numbers is converted into a binary system, it creates a 32-bit series. As in the next part of the procedure, the more sensitive least significant bits are selected to increase the randomness from these 32-bit series, and a bit series to be used in encryption is generated.

The bit series generated with consecutive numbers can not exceed the randomness tests. For this reason, bits in the appropriate order are selected from these 32-bit series. At the end of the RNG design, the NIST 800-22 [14] statistical test suite have been implemented to test the randomness of the generated bit sequence. The random bitstreams to be used as keys in encryption must pass all NIST tests. If the bit string generated here passes all NIST tests, this indicates that this bit sequence is the proper random feature to use in encryption.

In this design, if the state variables are observable, anyone can reproduce the same output bit sequence of the RNG and output of the encryption algorithm can be predicted. The authors of [8] have preferred to use NIST 800-22 [14] statistical test suite in order to analyze output randomness of their RNG design.

However, Big Crush [15] and Diehard [16] statistical test suites which are available at the publication date of target paper weren't applied to output bit stream of the RNG. It should be noted that, the target random number generation system [8] doesn't satisfy the first secrecy criteria, which states that "RNG must pass all the statistical tests of randomness."

3 Attack System

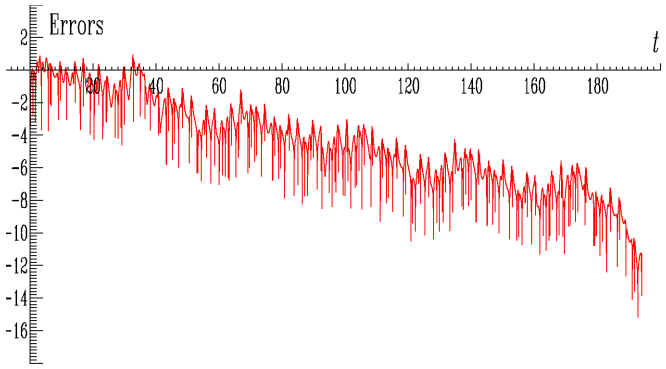


Fig. 2. Synchronization error Log $|e_x(t)|$.

After the seminal work on chaotic systems by Pecora and Carroll [18], synchronization of chaotic systems has been an increasingly active area of research [19]. In this paper, convergence of attack and target systems is numerically demonstrated using master-slave synchronization scheme [19]. In order to provide cryptanalysis of the target chaos-based encryption system an attack system is proposed which is given by the following Eqn. 2:

$$\begin{aligned} \dot{x_2} &= ay_2 - x_2 + z_2y_2 + k(x_1 - x_2) \\ \dot{y_2} &= -bx_2z_2 - cx_2 + y_2z_2 + d \\ \dot{z_2} &= e - fx_2y_2 - {x_2}^2 \end{aligned} \tag{2}$$

where k is the coupling strength between the target and attack systems. The only information available are the structure of the target chaos-based encryption system and a scalar time series observed from x_1.

In this study, we construct the attack system expressed by the Eqn. 2 that synchronizes ($x_2 \to x_1$ for $t \to \infty$) where t is the normalized time. We define the error signals as $e_x = x_1 - x_2$, $e_y = y_1 - y_2$, and $e_z = z_1 - z_2$ where the aim of the attack is to design the coupling strength such that $|e(t)| \to 0$ as $t \to \infty$.

The master-slave synchronization of attack and target systems is verified by the conditional Lyapunov Exponents (CLEs), and as firstly reported in [18], is achievable if the largest CLE is negative. Largest CLEs are calculated for different values of coupling strength k while a scalar time series is observable from x_1.

When k is greater than 0.44 and less than 1.60 the largest CLE is negative and hence identical synchronization of target and attack systems starting with different initial conditions is achieved and stable [18] (Largest conditional

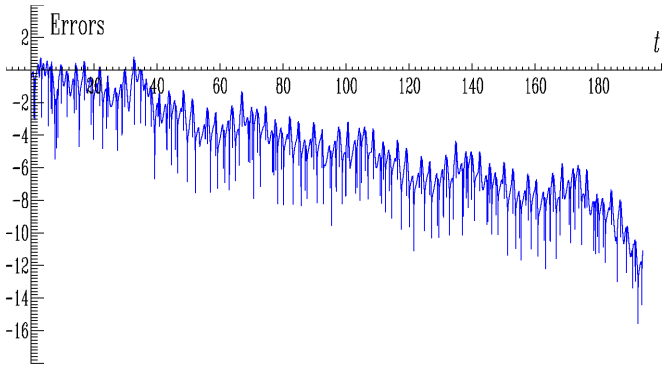


Fig. 3. Synchronization error Log $|e_y(t)|$.

Lyapunov Exponent is -0.000356331 for $k = 0.45$ and -0.0771864 for $k = 1.2$, respectively).

However for k is less than 0.44 or greater than 1.60, largest CLE is positive and identical synchronization is unstable (Largest conditional Lyapunov Exponent is 0.00521199 for $k = 0.4$ and 0.00118349 for $k = 1.6$, respectively).

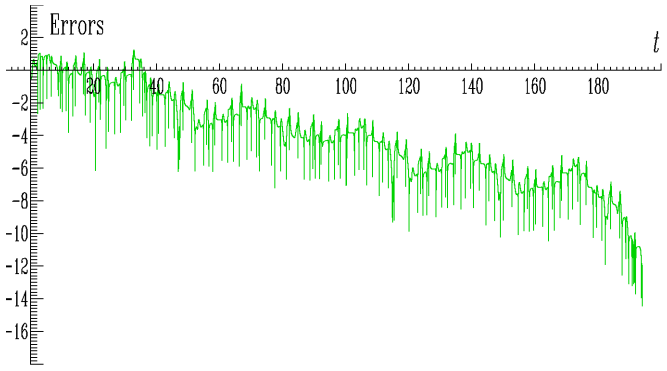


Fig. 4. Synchronization error Log $|e_z(t)|$.

Log $|e_x(t)|$, Log $|e_y(t)|$, and Log $|e_z(t)|$ are shown in Fig.2, Fig.3 and Fig.4 respectively, for $k = 1.2$, where the synchronization effect is better than that of $k = 0.45$. As shown in the given figures, the attack system converges to target system and master-slave synchronization is achieved in less than $190t$.

4 Numerical Results

We numerically demonstrate the proposed attack system using a 4^{th}-order Runge-Kutta algorithm with fixed step size and its convergence is illustrated in Fig.2, Fig.3 and Fig.4, respectively. Numerical results of $x_1 - x_2$, $y_1 - y_2$, and $z_1 - z_2$ are also given in Fig. 5, respectively illustrating the both synchronized and unsynchronized behaviors of target and attack systems.

It is observed from the given figures that, master-slave synchronization is achieved and stable. As shown by black lines in these figures, no synchronous phenomenon is observed before $190t$. In time, the proposed attack system converges to the target system and identical synchronization is achieved where colored lines depict synchronized behaviors of chaotic states in Fig. 5, respectively.

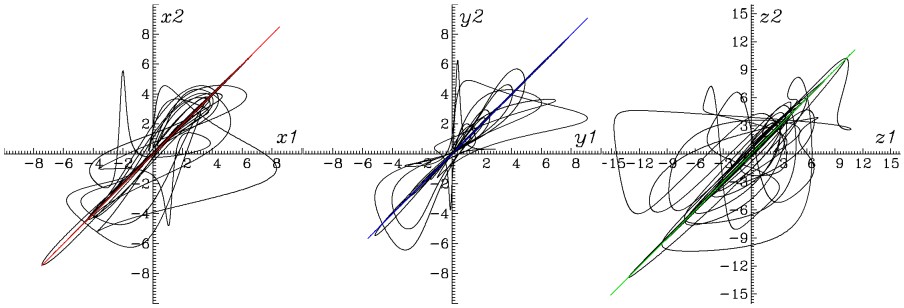


Fig. 5. Synchronized and unsynchronized behaviors of target and attack systems.

Since the identical synchronization of attack and target systems is achieved ($x_2 \rightarrow x_1$) in $190t$, the estimated values of output bit which is generated according to the procedure explained in Section 2 converge to their corresponding fixed values. As a result, it is obvious that identical synchronization of chaotic systems is achieved and hence output bit streams of target and attack RNGs are synchronized.

It is clearly shown master-slave synchronization of proposed attack system is achieved. Hence, output bit sequences of target and attack RNGs are synchronized. In conclusion, cryptanalysis of the target chaos-based encryption system not only predicts the previous and the next random bit but also demonstrates that the same output bit sequence of the target random number generation system can be reproduced. As a result, the target random number generation system [8] satisfies neither the second, nor the third secrecy criteria that a RNG must satisfy.

In [8], it was reported that the proposed chaos-based encryption algorithm uses less memory and offers higher encryption data rates compared with the traditional encryption algorithms. However, output of the chaos-based encryption algorithm can be predicted which is an unacceptable security weaknesses.

It should be noted that, deterministic chaos itself cannot be pointed out as the source of randomness.

5 Conclusions

In this paper, we propose a cryptanalysis method for a chaos-based encryption algorithm over TCP data packet. An attack system is introduced to discover the security weaknesses of the chaos-based encryption algorithm and its convergence is proved using master-slave synchronization scheme. Although the only information available are the structure of the chaos-based encryption algorithm and a scalar time series observed from the target chaotic system, identical synchronization of target and attack systems is achieved and hence output bit streams are synchronized. The target chaos-based random number generator (RNG) used for the encryption algorithm does not fulfill Big Crush and Diehard statistical test suites, the previous and the next bit can be predicted, while the same output bit sequence of the RNG can be reproduced. Simulation results presented in this paper not only verify the feasibility of the proposed method but also encourage its use for the cryptanalysis of the other chaos based encryption algorithms.

References

1. Devaney, R., An introduction to Chaotic Dynamical Systems, Second edition, Addison-Wesley, Reading MA, 1989.
2. Abel, A. and Schwarz, W. "Chaos communications principles, schemes, and system analysis", Proc. of IEEE, vol. 90, no. 5, pp. 691-710, May 2002.
3. Bucolo, M., Caponetto, R., Fortuna, L., Frasca M. and Rizzo, A., "Does chaos work better than noise?", IEEE Circuits and Systems Magazine, vol. 2, no. 3, pp. 4-19, 2002.
4. Small, M. and Tse, C. K., "Detecting determinism in time series: The method of surrogate data", IEEE Trans. Circuits Syst. I, vol. 50, pp. 663-672, May 2001.
5. Stojanovski, T., Kocarev, L. "Chaos-Based Random Number Generators-Part I: Analysis", IEEE Trans. Circuits and Systems I, Vol. 48, 3 (2001) 281-288
6. Callegari, S., Rovatti, R., Setti, G."Embeddable ADC-Based True Random Number Generator for Cryptographic Applications Exploiting Nonlinear Signal Processing and Chaos", IEEE Transactions on Signal Processing, Vol. 53, 2 (2005) 793-805
7. Stojanovski, T., Pihl, J. and Kocarev, L., "Chaos-Based Random Number Generators-Part II: Practical Realization," IEEE Transactions on Circuits and Systems I, vol. 48, no. 3, pp. 382-385, Mar. 2001.
8. Çavusoğlu, Ü., Akgül, A., Kaçar, S., Pehlivan, İ., and Zengin, A., "A novel chaos-based encryption algorithm over TCP data packet for secure communication," Security Comm. Networks, vol. 9 (2016), 1285 - 1296
9. Ergün, S., Güler, Ü., and Asada, K., "A High Speed IC Truly Random Number Generator Based on Chaotic Sampling of Regular Waveform" IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E94-A, no.1, (2011) 180-190

10. Ergün, S., Güler, Ü., and Asada, K., "IC Truly Random Number Generators Based on Regular & Chaotic Sampling of Chaotic Waveforms" Nonlinear Theory and Its Applications, IEICE transactions, vol. 2, no. 2, (2011) 246-261
11. Göv, N.C., Mıhçak, M.K. and Ergün, S.: True Random Number Generation Via Sampling From Flat Band-Limited Gaussian Processes. IEEE Trans. Circuits and Systems I, Vol. 58. 5 (2011) 1044-1051
12. Schneier, B.: Applied Cryptography. 2. edn. John Wiley & Sons (1996)
13. Menezes, A., Oorschot, P.van, Vanstone, S.: Handbook of Applied Cryptology. CRC Press (1996)
14. National Institute of Standard and Technology, FIPS PUB 140-2, Security Requirements for Cryptographic Modules, NIST, Gaithersburg, MD 20899, (2001)
15. L'Ecuyer, P., Universit'e de Montr'eal., "Empirical Testing of Random Number Generators", 2002, Available at http://www. iro.umontreal.ca/lecuyer/
16. Marsalgia, G., "Diehard: A Battery of Tests of Randomness", 1997, Available at http://stat.fsu.edu/∼geo/diehard.htm
17. National Institute of Standard and Technology, "A Statistical Test Suite for Random and Pseudo Random Number Generators for Cryptographic Applications", NIST Special Publication 800-22, Available at http://csrc.nist.gov/groups/ST/toolkit/rng
18. Pecora, L.M., Carroll, T.L., "Synchronization in chaotic systems," Physical Review Letters, vol. 64, no. 8, (1990) 821-824
19. Hasler, M., "Synchronization principles and applications," Tutorials IEEE International Symposium on Circuits and Systems (ISCAS '94), C. Toumazou, Ed., London, England, (1994) 314-327

Local fractal analysis of time series

E. Pardo-Igúzquiza (1), F. J. Rodríguez-Tovar (2) and J. Sanchez-Morales (2)

(1) Instituto Geológico y Minero de España, Ríos Rosas 23, 28003 Madrid (Spain)
(2) Universidad de Granada, Campus Fuentenueva s/n, 18071 Granada (Spain)

e.pardo@igme.es; fjrtovar@ugr.es; josesanmor@correo.ugr.es

Abstract. In geosciences, in addition to proper time series there are one-dimensional data sequences where time is substituted by space. That is the case, for example, with data measured along a borehole or along a transect in the field or on a map. In all cases, the fractal dimension of the series is of interest in order to identify variability dependence on scale. However, if the entire time series is used in the analysis, the monofractal model may not be appropriate due to multifractality. On the other hand, the monofractal model is more likely to be adequate for a smaller range of scales that can be analyzed inside a moving window running along the sequence of data. A variogram is proposed for local fractal dimension estimation and the result is a new sequence with the local fractal dimension. The new time series can be used in different ways, such as for signal segmentation and classification.

1 Introduction

Fractal analysis of time series has different applications in science [1]. Broadly speaking, the fractal dimension of a time series is related to its roughness (or smoothness) [2]. In geosciences, in addition to proper time series there are one-dimensional data sequences where time is substituted by space. That is the case with data measured along a borehole, a stratigraphic sequence, or data taken along transects. Furthermore, in geosciences, fractal analysis along transects is used in the fractal analysis of surfaces [3]. In any event, fractal analysis is used across disciplines in medicine [4], econometrics [5], engineering [6], and Internet analysis [7].

There are many methods that have been proposed for fractal analysis of time series. That is, there are many estimators of fractal dimension. Among these methods, the box-count, divider, level crossing, variogram, madogram, power spectra, and wavelet methods [2] are some of the most popular. In this work we have chosen the variogram because it is especially suited for fractal analysis of short sequences and because it

adfa, p. 1, 2011.

has been chosen as one of the best methods in comparison studies [2], [3]. The fractal analysis is made local by using an overlapping moving window as described in the next section.

2 Methodology

A time series (or a space sequence or transect) of n data can be represented as$\{X_1, X_2, \dots, X_n\}$, where X_t is a variable measured at time t or at spatial location t. In fractal analysis in geosciences, the variogram method is based on the observation that a fractional Brownian motion (fBm) is a good stochastic model for spatial variables [8] and its variogram [9] has a power-law model [10]:

$$\gamma(k) = \frac{1}{2}\mathrm{E}[X_{t+k} - X_t]^2 = \alpha k^{\beta} \tag{1}$$

$$\alpha \geq 0; 0 \leq \beta < 2,$$

where t are the coordinates (time or spatial) of a sampled location, k is a vector of time or space distance, $\gamma(k)$ is the variogram for a distance vector k, E{.} is the mathematical expectation operator and α and β are scaling parameters. The fractal dimension D is related to the scaling parameter β [11] by:

$$D = E + 1 - \frac{\beta}{2}, \tag{2}$$

where E is the topological dimension of the Euclidean space in which the fBm is distributed. For a one-dimensional fBm (i.e. a time series, a profile or transect, etc.), $E = 1$, and:

$$D = 2 - \frac{\beta}{2}. \tag{3}$$

Taking logarithms of both sides of equation (1) results in:

$$\log[\gamma(k)] = \beta \cdot \log[k] + \log[\alpha] \tag{4}$$

Thus, β can be estimated from the slope of a line fitted to the log-log plot of the variogram and the fractal dimension can be estimated from equation (3).

The variogram is estimated for a number of lags equal to K. That is, there are K data pairs for fitting a straight line in equation (4). Using standard normal regression, slope β may be estimated, as well as its standard error (SE(β)) and the coefficient of determination (R^2) of the fit [12].

Local fractal analysis implies calculating the fractal dimension inside a moving window of size $(2W + 1)$, where W is half the size of the window. Thus, the result is a new time series of fractal dimensions, $\{D_1, D_2, \dots, D_n\}$, where each fractal dimension

D_t is calculated by the variogram method using the data inside the window centered at time t [13]:

$$D_t = FD[X_{t-W}, X_{t-W+1}, \dots, X_{t-1}, X_t, X_{t+1}, \dots, X_{t+W-1}, X_{t+W}], \quad (5)$$

Here, $FD[.]$ is the operator that estimates the fractal dimension, that is, the statistical estimator of the fractal dimension. The methodology is illustrated in the next section.

3 Case study

The digital elevation model (DEM) describes topography as a matrix of numbers that represent the elevations of the terrain. DEM are often used in the context of Geographical Information Systems (GIS) [14]. The entire country of Spain is available at a resolution of 5 m (i.e. each pixel represents the mean altitude of the terrain in a square measuring 5 m by 5 m in terrain units) and can be downloaded for free from the web page of the Instituto Geográfico Nacional [National Geographic Institute]. For example, Fig. 1 shows the DEM of the Sierra de las Nieves karst aquifer [15], which has 4600 columns and 3600 rows. A profile has been selected, as indicated by a black line in Fig. 1, and has been represented in Fig. 2. The terrain transect of Fig. 2 is no different from an evenly sampled time series, where the variable measured is altitude of the terrain in meters above sea level. Fig. 2 (top) represents the full sequence, where the smoothness of the sequence changes along the transect. Fig. 2 (bottom) shows a detail of Fig. 2 (top) in order to illustrate the concept of statistical self-similarity and to show how a section that appears smooth at a given scale is really rough when magnified. Thus, a local fractal analysis will provide the variation of roughness along the sequence.

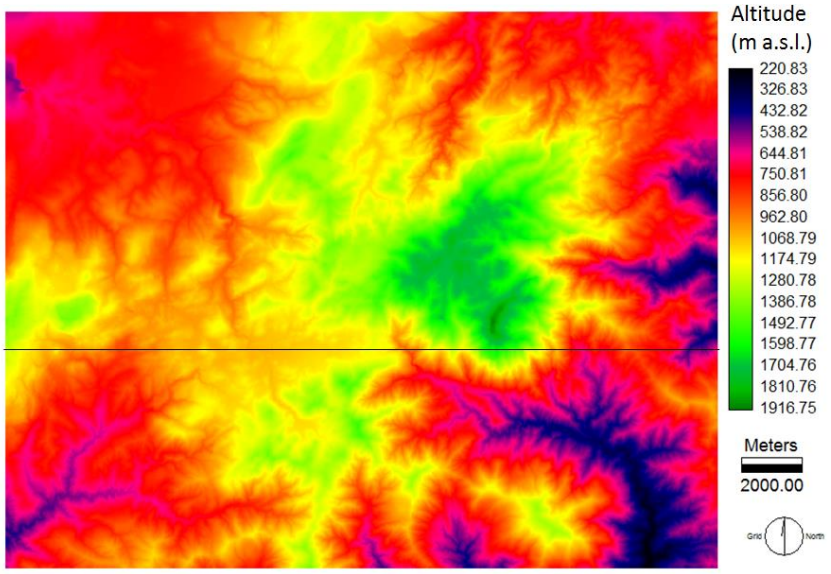
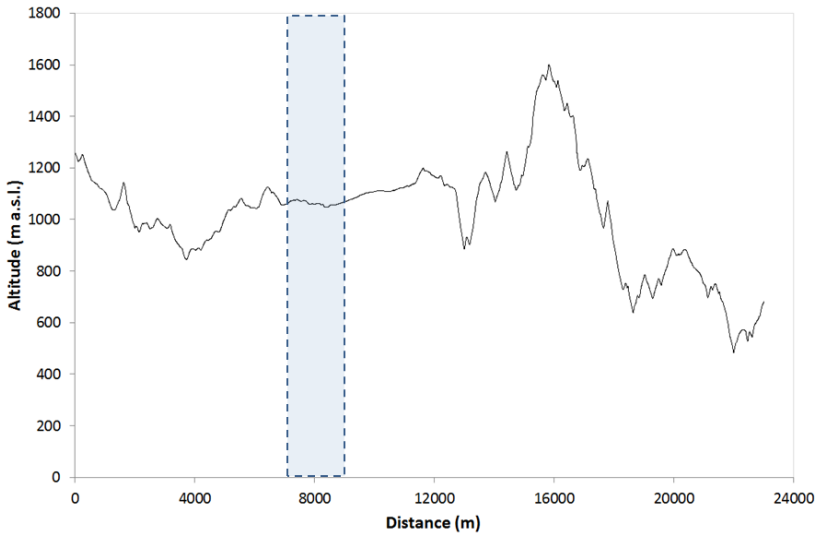


Fig. 1. Digital elevation model with 4600 columns and 3600 rows and with a spatial resolution of 5m. The solid black line is the chosen transect represented in Fig. 2.



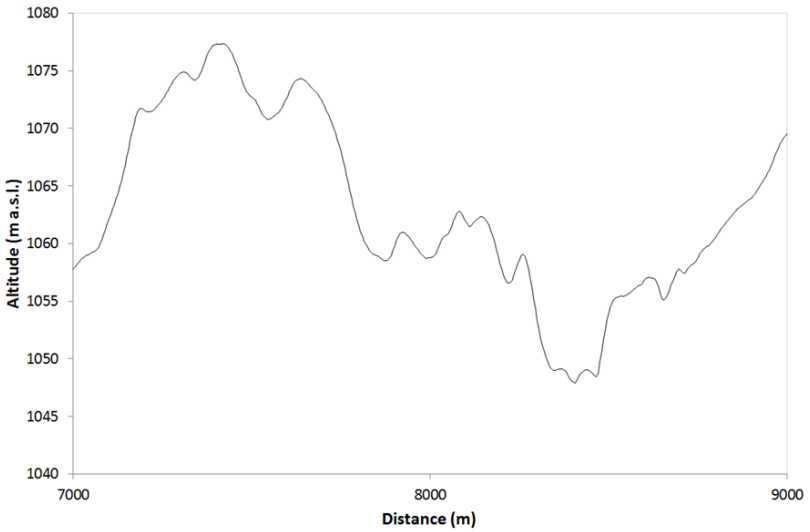


Fig. 2. Top: Topography transect represented in Fig. 1 by the solid black line. Bottom: Detail of the transect between coordinates 7000 and 9000, which is represented by the shaded rectangle in the top figure.

By calculating the fractal dimension of the entire sequence using the variogram method, a fractal dimension of 1.1 has been obtained, as shown in Fig. 3. There are several problems when calculating the fractal analysis of the entire sequence. First, the overall fractal dimension averages the fractal variability along the sequence as a single number and therefore smooths the fractal dimension, which may actually be different at different locations along the transect. Secondly, the graph in Fig. 3 begins to show a curvature for large distances, which implies that the mono-fractal model is only valid for a scaling range due to the multifractal nature of topography [16]. The local fractal analysis applies fractal analysis inside a moving window which contains a given scaling range and thus the monofractal model can be applied to obtain the fractal dimension along the sequence. This result is shown in Fig. 4, where it can be seen that the fractal dimension is highly variable, ranging from 1 to 1.56, and with a mean of 1.12. The standard error of the fractal dimension is shown in Fig. 5, and it ranges from 0.019 to 0.0038, with a mean value of 0.032. Thus, the fractal dimension is estimated with high confidence, with a degree of reliability that extends to the second decimal place. Furthermore, Fig. 6 shows the coefficient of determination (R^2), which quantifies the goodness of fit of a straight line to the regression equation. Its mean value is 0.988 and the 10th percentile is 0.972, meaning that the monofractal model is a good approximation for the range of scales used in the local fractal analysis of the topographic profile. The moving window had a size of 405 meters (81 cells) and the number of variogram lags used in the analysis was 30.

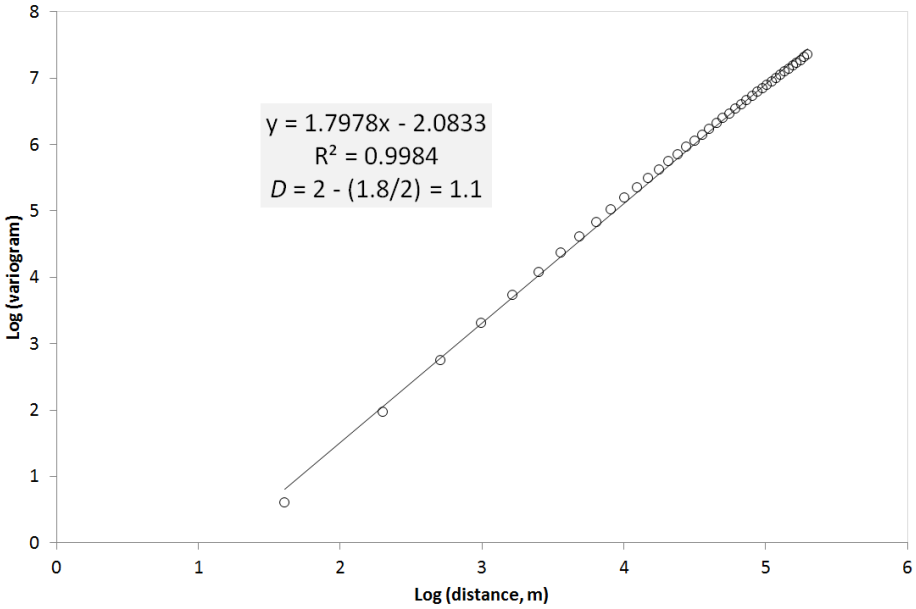


Fig. 3. Fractal dimension (*D*) of the entire transect in Fig. 2.

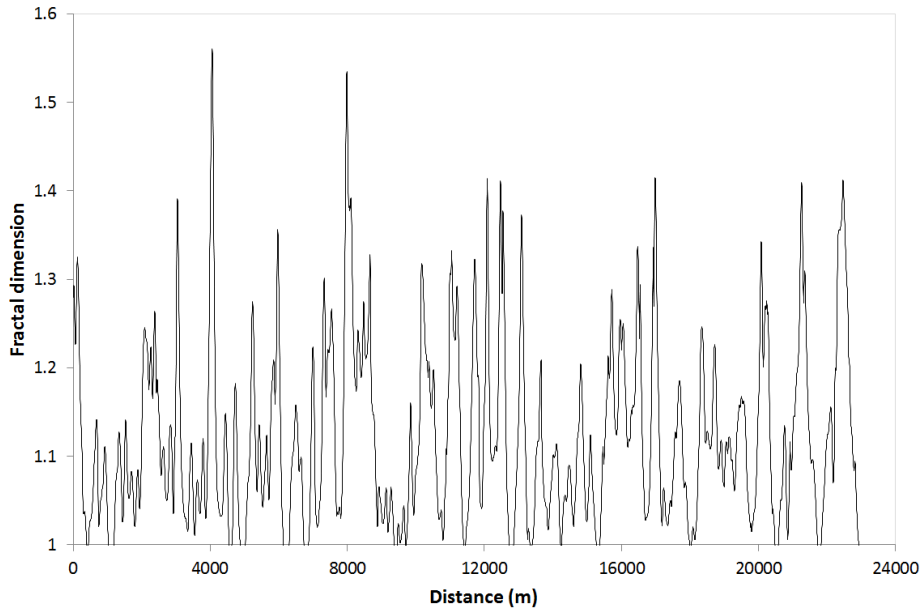


Fig. 4. Local fractal dimension along the transect in Fig. 2.

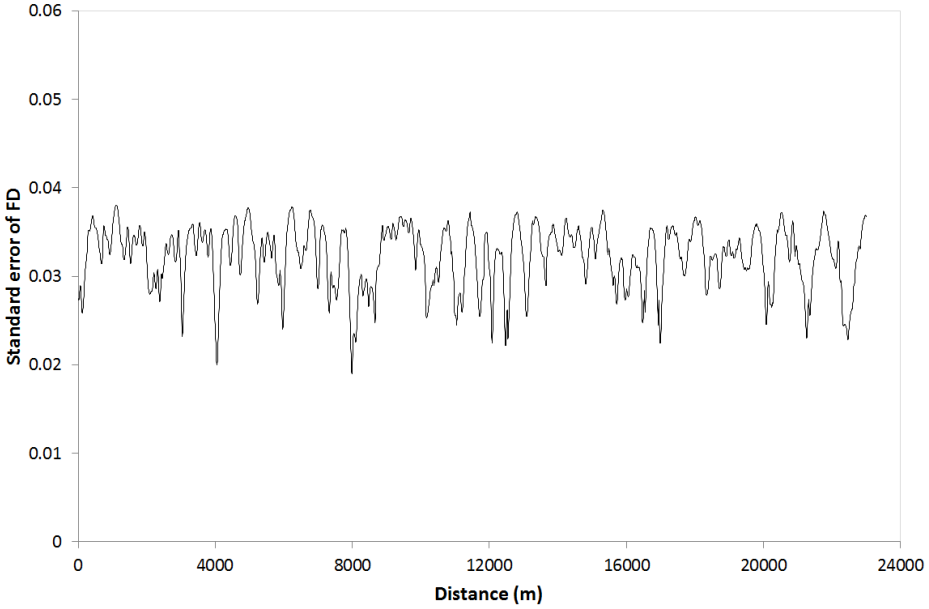


Fig. 5. Standard error of the local fractal dimension in Fig. 4.

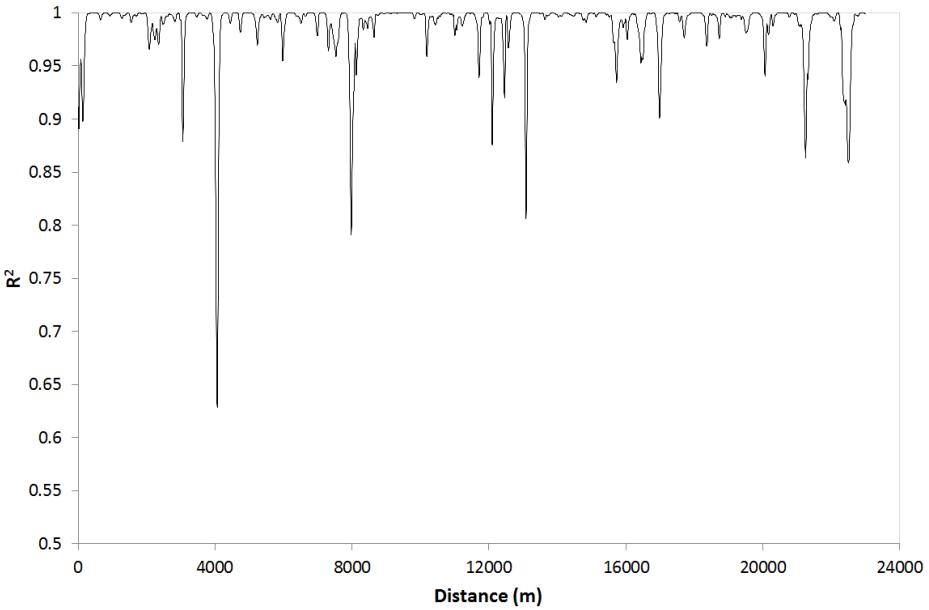


Fig. 6. Coefficient of determination (R^2) of the estimation of fractal dimension along the transect of Fig. 2.

The local fractal dimension can be used in signal segmentation. This has been done by using a fractal dimension threshold of 1.15 and classifying the topographic transect into two kinds of terrain according to its local fractal dimension. The new sequence is shown in Fig. 7. If the local fractal dimension is larger than 1.15, the terrain shows a high local fractal dimension and is classified as rough (with a classification value of 2 in Fig. 7), and if the local fractal dimension is below 1.15, the terrain is classified as smooth (value 1 in Fig. 7). This method can be used to identify areas of interest in the field that could be flagged for future visits in order to understand the origins of such irregularities, which in the case study area are related to karst geoforms [17].

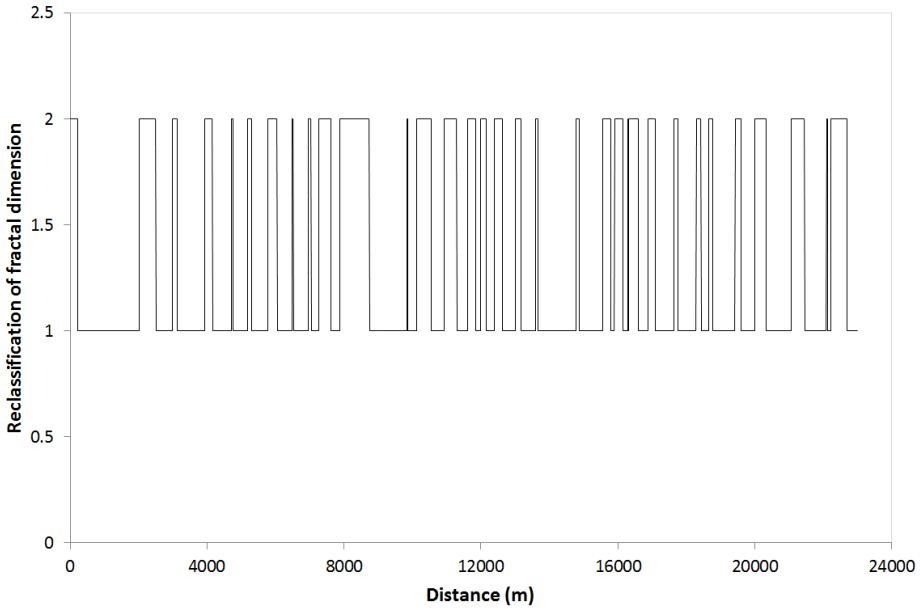


Fig. 7. Reclassification of fractal dimension along the transect using the threshold of 1.15 and creating a new sequence with low fractal dimension segments (1) and high fractal dimension segments (2) along the transect.

4 Conclusions

Local fractal analysis is a new tool for analyzing the roughness (smoothness) of time series and space transects; its main applications are in signal segmentation and classification, among others. This method has been illustrated with a transect of topography. In this case, local fractal analysis allows for the use of a monofractal model with a variable which is multifractal in order to identify areas of special geological interest. Further applications could be identified by applying local fractal analysis to other time series or spatial sequences.

Acknowledgements

This work has been supported by the research project CGL2015-71510-R from the Ministerio de Economia, Industria y Competitividad.

References

1. Mandelbrot, B. B., 1982. The fractal geometry of Nature. W. H. Freeman and Co., San Francisco, CA.
2. Gneiting T, Ševčíková H, Percival DB. 2012. Estimators of fractal dimension: assessing the roughness of time series and spatial data. Statistical Science 27: 247-277.
3. Klinkenberg B, Goodchild MF. 1992. The fractal properties of topography: a comparison of methods. Earth Surface Processes and Landforms 17: 217-234.
4. Klonowski W., 2016. Fractal Analysis of Electroencephalographic Time Series (EEG Signals). In: Di Ieva A. (eds) The Fractal Geometry of the Brain. Springer Series in Computational Neuroscience. Springer, New York, NY
5. Kapecka, A., 2013. Fractal Analysis of Financial Time Series Using Fractal Dimension and Pointwise Hölder Exponents. Dynamic Econometric Models, 13, 107-125.
6. Zhang, X-y. and Luo, L-y., 2012. Self-similarity analysis of time series IEEE 11th International Conference on . Signal Processing (ICSP).
7. Stathis, C., and Maglaris, B., 2000. Modelling the self-similar behaviour of network traffic. Computer Networks, 34(1), 37-47.
8. Arakawa K, Krotkov E, 1996. Fractal modeling of natural terrain: analysis and surface reconstruction with range data. Graphical Models and Image Processing 58: 413-436.
9. Matheron G. 1963. Principles of geostatistics. Economic Geology (1963) 58: 1246-1266.
10. Wen R, Sinding-Larsen R. 1997. Uncertainty in fractal dimension estimated from power spectra and variograms. Mathematical Geology 29: 727-753.
11. Voss RF. 1985. Random fractal forgeries, in Earnshaw, R.A., ed. Fundamental algorithms for computer graphics: NATO ASI Series, v. F17, 805-835.
12. Norman R. Draper, N. R. and Smith, H., 1998. Applied Regression Analysis. John Wiley & Sons, Inc., 706 p.
13. Chaudhuri, B. B. and Sarkar, N., 1995. Texture segmentation using fractal dimension. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17, 72-77.
14. Burrough, P. A., McDonnell, R. and Lloyd, C. D. 2015. Priciples of Geographical Information Systems, Oxford University Press, Third Edition, Oxford (UK).
15. Pardo-Igúzquiza E., Durán-Valsero J.J., Dowd P.A., Guardiola-Albert C., Liñan-Baena C. and Robledo-Ardila P.A., 2012. Estimation of spatio-temporal recharge of aquifers in mountainous karst terrains: application to Sierra de las Nieves (Spain). Journal of Hydrology, 470-471, 124-137.
16. Gagnon, J.-S., Lovejoy, S. and Schertzer, D., 2006. Multifractal earth topography. Nonlinear Processes in Geophysics, 13, 541-570.
17. Pardo-Igúzquiza E., Durán J.J. and Dowd P.A. 2013. Automatic detection and delineation of karst terrain depressions and its application in geomorphological mapping and morphometric analysis. Acta Carsologica 42/1, 17-24.

Discussion on Geodetic Times Series of Mixed Spectra and Levy Processes

J.-P. Montillet and K. Yu

Formerly at Ecole Polytechnique Federal de Lausanne, Switzerland
Adjunct Researcher at Space and Earth Geodetic Analysis Laboratory (SEGAL),
UBI/IDL, Portugal. {jeanfi_montillet}@yahoo.fr
School of Geodesy and Geomatics, Wuhan University, China
{kgyu}@sgg.whu.edu.cn

Abstract. Recently, various models have been developed, including the fractional Brownian motion (fBm), to analyse the stochastic properties of geodetic time series, together with the extraction of geophysical signals. The noise spectrum of these time series is generally modelled as a mixed spectrum, with a sum of white and coloured noise. Here, we are interested in modelling the residuals, after deterministically subtracting geophysical signals from the observations. This residual time series is then assumed to be a sum of three random variables (r.v.), with the last r.v. belonging to the family of Levy processes. This stochastic term models the remaining residual signals and other correlated processes. We identify three classes of Levy processes: Gaussian, fractional and stable. In the first case, residuals are predominantly constituted of short-memory processes. Fractional Levy process can be an alternative model to the fBm in the presence of long-term correlations and self-similarity property. Stable processes have infinite variance, which can mainly be satisfied in the case of heavy-tailed distributions in the application to geodetic time series, and then implying an important misfit in the modelling of both the underlying processes and geophysical signals. Finally, we support that the use of the α-stable distributions with geodetic time series can only be justified in the particular case of the stable Levy processes.

Keywords: Levy Processes, Geodetic Time Series, α Stable Distribution, Mixed Spectra, Heavy Tails, fBm, Coloured Noise, Gaussian Increments, Hurst Parameter, ARMA, FARIMA

1 Introduction

Among the geodetic data, time series of daily position of Global Navigation Satellite System (GNSS) receiver have been of particular interest for the study of geophysical phenomenon at regional and global scales (e.g., study of the crustal deformation due to large Earthquakes, sea-level rise -(Bock and Melgar , 2016; Herring et al. , 2016)). However, this data contains white noise and long-memory processes (i.e. coloured noise). The scientific community agrees with the existence of a trade-off in estimating both the stochastic and functional models (He et al.

, 2017). More precisely, the choice of the stochastic model directly influences the estimation of the geophysical signals included in the functional model (i.e., tectonic rate, seasonal variations, slow-slip event - (Bock and Melgar , 2016; He et al. , 2017)). To name a few, it includes the First Order Gauss-Markov (FOGM) model, the white noise with power-law noise (Williams , 2003; Williams et al. , 2004), the Generalized Gauss Markov noise model (GGM), or the Band-pass noise (Langbein , 2008). The optimal selection of the stochastic model in GNSS time series analysis remains a hot topic in the scientific community (Bock and Melgar , 2016; Herring et al. , 2016; He et al. , 2017).

Recently, Montillet and Yu (2015) proposed the use of the fBm, first developed by Mandelbrodt et al. (1968), in order to model long-memory processes. Botai et al. (2011) and Montillet and Yu (2015) focused on modelling (residual) geodetic time series using the family of Levy α-stable distributions (Nolan , 2009). The application of this family of distribution was supported by the ability to model long-memory processes and the existence of impulsive signals/noise bursts in the data sets suggesting deviations from Gaussian distribution (Botai et al. , 2011).

Here, we discuss several statistical assumptions (i.e. stationary properties, presence of long-term correlations, Gaussianity of the increments) on the underlying processes in the residual GNSS time series. We focus on justifying the application of the fractional Brownian motion (fBm) and the family of Levy α-stable distributions in the work of Montillet and Yu (2015).

The next section starts with the definition of the residual geodetic time series, the fBm and the relationship with the Fractional Autoregressive Integrated Moving Average (FARIMA) model. From financial analysis, we introduce the family of Levy processes (Panas , 2001) and the assumptions in order to relate to other models (i.e. FARIMA, fBm). Section 3 presents the assumptions and a discussion on the use of the Levy processes in the model of the residual time series. To do so, we model the residual geodetic time series as a sum of three random variables, with the hypothesis that the third one is a Levy process. It involves some justifications compared with established models in the scientific community. This section ends with the limits of modelling geodetic time series with Levy processes.

2 The Stochastic Properties of the Residual Time Series and the Definition of Levy Processes

2.1 Model of Residual GNSS Time Series

GNSS time series are generally regarded as a sum of geophysical signals (i.e., seasonal signal, tectonic rate) and stochastic processes (e.g., white noise, coloured noise) (Williams et al. , 2004; Davis et al. , 2012). Here, the residual time series are defined as the remaining time series after subtracting deterministically modeled tectonic rate and seasonal components, i.e. the functional model, from the GNSS observations. The functional model of those signals is based on the

polynomial trigonometric method (Li et al. , 2000; Williams , 2003; Tregoning and Watson , 2009)

$$s_0(t) = \sum_{j=1}^{N}(c_j \cos(d_j t) + e_j \sin(d_j t)) + at + b \tag{1}$$

with $s_0(t)$ the sum of the seasonal signal and tectonic rate at the epoch t. Note that d_j is equal to $2\pi j/N$, and N equal to seven (He et al. , 2017). If $x(t)$ is the residual time series after subtracting the GNSS observations with the functional model (i.e. $s_0(t)$), one can then formulate the hypothesis that the residual time series is a sum of a residual signal and a noise. We can describe the stochastic noise model with the variance of the vector of noise components $\mathbf{n} = [n(t_1), n(t_2), ..., n(t_L)]$, such as (Williams , 2003):

$$\sigma_n^2(L) = E\{\mathbf{n}^\dagger \mathbf{n}\} = \sigma_{n0}^2 \mathbf{I} + \sigma_{n1}^2 \mathbf{J} \tag{2}$$

$\dagger$ is the transpose operator, $\mathbf{I}$ is the identity matrix, $\mathbf{J}$ is the variance-covariance matrix of the coloured noise, σ_{n0}^2 and σ_{n1}^2 are the variance of the white noise and coloured noise respectively. Therefore, this type of time series belongs to the family of mixed spectra, where the mixed spectrum results from the sum of the spectra corresponding to the two kinds of noise (Li , 2013). Note that the length of the time series L is much larger than the number of frequencies N defining $s_0(t)$. In addition, we assume the Gaussian noise to be zero-mean, whereas the coloured noise with a mean equal to μ_C, slowly varying with the length of the time series. That is to satisfy the wide-sense stationary hypothesis (Kasdin , 1995; Haykin , 2002). Finally, the residual signal is considered to be the remaining geophysical signals (i.e. seaosnal component and tectonic rate) not completely estimated due to the mismodelling of the stochastic properties of the time series and other small amplitude (sub millimeter) signals (i.e. co-seismic offsets, transients) (Gazeaux et al. , 2013; Herring et al. , 2016; He et al. , 2017).

2.2 Modelling the Stochastic Properties with fBm, FARIMA and Power-law

The error spectrum of the GNSS time series is best characterised by a stochastic process following a power-law with index β ($0 \leq \beta \leq 2$) (Herring et al. , 2016; He et al. , 2017). A power-law noise model means that the frequency spectrum is not flat but is governed by long-range dependencies. If the probability density function of the noise is Gaussian or has a different density function with a finite value of variance, its fractal properties can be described by the Hurst parameter (H). Montillet et al. (2013) proposed to use the fBm in order to model the statistical properties of the residual time series. Following the definition from Mandelbrodt et al. (1968), if $H < 0.5$, the process behaves as a Gaussian variable (anti-persistent); if $H > 0.5$ the process exhibits long-range dependence (persistent); while the case of H equal to 0.5 corresponds to a pure Brownian motion (white noise). The same study showed that H is directly connected with

β by the relation $\beta = 2H - 1$.

Long-memory processes are modeled with a specific class of Autoregressive Moving Average (ARMA) models, the so-called FARIMA (or ARFIMA in Montillet and Yu (2015)), by allowing for non-integer differentiating. A comprehensive literature on the application of this model can be found in financial analysis (Panas , 2001), and in computer science (Contreras-Reyes and Palma , 2013). A time series $x(t)$ follows an FARIMA (p,d,q) process (with p and q in $\mathbb{Z}^+$, and d in $\mathbb{R}^+$) if it can be defined by

$$\begin{aligned} \Psi_p(Z)x(t) &= \Theta_k(Z)(1-Z)^{-d}b(t) \\ \Psi_p(Z) &= 1 - a_1 Z - a_2 Z^2 - ... - a_p Z^p \\ \Theta_k(Z) &= 1 + b_1 Z + b_2 Z^2 + ... + b_k Z^k \end{aligned} \tag{3}$$

where $[b_1, b_2, ..., b_k]$ are the MA parameters and $[a_1, ..., a_p]$ the AR parameters, with $\sum_{i=1}^{k} b_i^2 < \infty$. The properties of the FARIMA model are presented by Granger and Joyeux (1980): i) if the roots of the polynomial functions $\Psi_p(Z)$ and $\Theta_k(Z)$ are outside the unit circle and $d < |0.5|$, then $x(t)$ is both stationary and invertible; ii) if $0 < d < 0.5$ the FARIMA model is capable of generating stationary series which are persistent. In this case the process displays long-memory characteristics, with a hyperbolic autocorrelation decay to zero; iii) if $d \geq 0.5$ the process is non-stationary ; iv) when d equal to 0 it is an ARMA process exhibiting short memory; v) when $-0.5 \leq d < 0$ the FARIMA process is said to exhibit intermediate memory or anti-persistence. This is very similar to the description of the Hurst parameter in the fBm model. In fact, d and H are linked by the relationship $H = d + 0.5$ (Panas , 2001).

2.3 Definition of the Levy Processes and the Relationship with fBm in Time Series Analysis

The fractional Levy distribution models the Levy processes and in particular the general family of α-stable Levy processes used in financial time series analysis (Panas , 2001; Wooldridge , 2010). Let us recall the definition of a stable random variable.

Definition *Nolan (2009), chap. 1, definition, 1.6* A random variable X is stable if and only if $X \stackrel{d}{=} aZ + b$, where $a \neq 0$, $b \in \mathbb{R}$ and Z is a random variable with characteristic function $\phi(u)$, which is the Fourier transform of the probability density function (Nolan , 2009):

$$\phi(u) = \begin{cases} \exp\left(j\mu u - \gamma|u|^{\alpha}[1 - jksign(u)\tan(\frac{\pi\alpha}{2})]\right) if \ \ \alpha \neq 1 \\ \exp\left(j\mu u - \gamma|u|[1 - j\frac{2}{\pi}ksign(u)log(|u|)]\right) if \ \ \alpha = 1 \end{cases} \tag{4}$$

Where *sign* is the signum function, α ($0 < \alpha \leq 2$) is the characteristic exponent which measures the thickness of the tails of the distribution (the smaller the values of α, the thicker the tails of distribution are), $k \in [-1, 1]$ is the symmetry

parameter which set the skewness of the distribution, γ is the positive scale parameter, and $\mu \in \mathbb{R}$ is the location parameter. In general, no closed-form expression exists for this distribution, except for the Gaussian (α equal to 2), Pearson (α equal to 0.5, k equal to -1) and Cauchy (α equal to 1, k equal to 0) distributions. Note that the distribution is called a symmetric α-stable if $k = \mu = 0$ (Wang et al. , 2008; Nolan , 2009). In order to estimate the four parameters, various methods exist such as maximum-likelihood (Nikias and Shao , 1995), regression numerical method (Koutrouvelis , 1980) and numerical integral method (Nolan , 2009).

$\phi(u)$ is defined for the random variable Z such as $\phi(u) = E\{\exp(iuZ)\} = \int_{-\infty}^{\infty} \exp(iuz)F(z)dz$. $F(z)$ is the distribution function of Z. $E\{.\}$ is the expectation operator. When the α-stable distributions are symmetric around zero, the characteristic function of aZ has the simpler form:

$$\phi(u) = \exp(-a^{\alpha}|u|^{\alpha}) \tag{5}$$

Therefore, the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ is stable with $\alpha = 2$, $k = 0$, $a = \sqrt{\sigma}/2$, $b = \mu$ (Nolan , 2009).

If a stochastic process is self-similar, then one can model it with a fBm model (see (Cont and Tankov , 2004), Definition 7.1). Recently, various works (Benassi et al. , 2004; Marquadt , 2006) have shown that fBm can be generalized by (Fractional) Levy processes under specific conditions (i.e. finite variation of the path, real harmonisable, well-balanced - (Benassi et al. , 2004)). The technical details of this generalization can be found in (Marquadt , 2006), but it is outside the scope of our work. Here, we define a Fractional Levy process by integral transformations replacing the Brownian motion by more general Levy process in the integral representation of fBm (Mandelbrodt et al. , 1968).

Moreover, there is a relationship between the (symmetric) α-stable Levy process and the Hurst parameter (H). A self-similar Levy process is either Brownian ($H = 1/2$) or must have infinite variance (Cont and Tankov , 2004). For $H \geq 1/2$, a self-similar Levy process is a (symmetric) α-stable Levy process (Cont and Tankov , 2004) (p. 236). Further important properties of the stable random variable are not discussed here, but reader can refer to Cont and Tankov (2004) and Nolan (2009). Finally, this family of distributions is of a particular interest in this work as the α index is equal to the inverse of the Hurst parameter. Panas (2001) stated that for $1/\alpha < H$, the stochastic processes exhibit persistence or positive increments; whereas for $0 < H < 1/\alpha$ the stochastic properties exhibit anti-persistence or negative increments. As a consequence, this family of distributions should be better suited when modeling the residual time series with a strong coloured noise or including long-memory processes; whereas if the white noise is predominant, the time series should be fitted with a Gaussian distribution.

3 Application of Levy Processes in Modelling Geodetic Time Series

3.1 Modelling Time Series of Mixed Spectra with Levy Processes

In order to investigate the application of Levy processes in modelling the residual GNSS time series, we look at the residual as a sum of three random variables. Namely, it is the sum of a white noise, a coloured noise and a third random variable (r.v.). It is a comparable approach used previously in Langbein (2012) and Davis et al. (2012). The stochastic properties of the third r.v. should tell us how well fits the selected models (i.e., functional and stochastic) to the observations. To recall the definition of the Levy processes in Section 2.3, we postulate that the third r.v. belongs to the Levy process. We then list the type of Levy processes (Wooldridge , 2010; Cont and Tankov , 2004) depending on the assumptions on the underlying stochastic process:

1- (Gaussian Levy) The Levy process is a Gaussian Levy process if the r.v. follows the properties of a Wiener process (identity matrix of covariance, zero-mean, stationary process - (Haykin , 2002; Wooldridge , 2010)). The residual time series is assumed to contain only short-term correlation processes modeled with an ARMA process.
2- (Fractional Levy) The residual time series has Gaussian increments and exhibits self-similarity with possibly long-term correlation processes. The Fractional Levy process models short and long-term correlations similar to the fBm (Cont and Tankov , 2004). The long-term correlation processes can be due to the presence of high-amplitude coloured noise (He et al. , 2017). As explained in Montillet and Yu (2015), the ratio of the amplitude of the coloured over white noise determines which stochastic model of the residual time series should be the most suitable between the FARIMA and ARMA processes.
3- (Stable Levy) The Levy process is a (α) stable Levy process. That is to generalize important misfit between the selected (stochastic and functional) model ($s_0(t)$) and the observations. If small jumps are presents and cannot be detected, the Levy process can be a Poisson Levy process to model those jumps as Markov jumps (Cont and Tankov , 2004). The result is that the distribution of the residual time series can be severely skewed, not symmetric, with the apparition of heavy tails, hence modeling with a Levy α-stable distribution. With the relationship between α, H and d (see Section 2.3), the stochastic properties of the residual time series should be best described with the FARIMA process.

The assumption of modelling jumps as Markov jumps in the residual GNSS time series may not be intuitive, because the general model is a Heaviside step function (Gazeaux et al. , 2013; Herring et al. , 2016; He et al. , 2017). Those jumps result from equipment changes (i.e. antenna, radom) to the receiver, sudden events (bumps to the antenna), geophysical nature (co-seismic offsets) and variations in the environment of the receiver occasioning multipath (growing trees, buildings).

In financial time series, the jumps are generally resulting from the randomness of the stock prices and modelled as random-walk. That is why in order to assume a (stable) Levy processes as the underlying stochastic model in geodetic time series, we need to restrict our assumption to small jumps not detectable by the human eyes, modelling them potentially as random-walk (Gazeaux et al. , 2013).

Now, one can underline that the sum of the three r.v. should follow the central limit theorem (CLT) for the Levy Gaussian and fractional Levy processes in the absence of large tails and finite variance (Papoulis and Unnikrishna Pillai , 2002) (*Chapter 7, p 278*). However, the stable Levy process should follow the generalized CLT (Nolan , 2009)(*Chap.* 1, *Theorem* 1.20) due to the properties of the (α) stable Levy process. Also, very large variance characterizing α-stable processes, is, in our assumptions, due to residual geophysical signals (i.e. residual tectonic rate or seasonal signal, large offsets, outliers). Therefore, those features can be detected a priori by a careful examination of the time series.

3.2 Limits of Modeling with Levy Processes

As discussed in the previous sections, the stable Levy process is characterized by a very large (or infinite) variance. In Montillet and Yu (2015), it was assumed that the infinite variance of the residual time series comes from large tails of the distribution (also called heavy tails -(Wooldridge , 2010)), generated by a large amplitude of coloured noise, outliers and other remaining geophysical signals. The same study implied that the values of the noise variance should be bounded, excluding extreme values. This is an important assumption to decide whether or not (symmetric) α-stable distributions can be used to model any geodetic time series. Here, we are investigating how the variance due to residual tectonic rate or seasonal signal evolves with the length of the residual time series (i.e. L epochs).

To recall Section 2.1 and the assumption on the noise properties, let us estimate the mean $< x(L) >$ and variance σ_x^2 of the residual time series (x) over the L epochs (i.e. considering the L-th epoch defined as $t_L = Ldt$, with the sampling time dt equal 1 for simplification and without taking into account any missing epoch in order to have a continuous time series). Based on Papoulis and Unnikrishna Pillai (2002), one can estimate the variance σ_x^2 in the cases of remaining linear trend, such as:

$$
\begin{aligned}
x(t_i) &= a_r t_i + b_r + n(t_i) \\
< x(L) > &= \frac{1}{L}\sum_{i=1}^{L}(a_r t_i + b_r + n(t_i)) \\
< x(L) > &= b_r + a_r \frac{L(L+1)}{2} + \mu_C \\
< x(L) > &\simeq a_r \frac{L^2}{2}
\end{aligned}
$$

$$\sigma_x^2(L) = \frac{1}{L}\sum_{i=1}^{L}(x(t_i) - < x(L) >)^2$$
$$\sigma_x^2(L) = \frac{a_r^2(L+1)(2L+1)}{6} - \frac{a_r^2(L+1)^2}{4} + \sigma_n^2(L) - \mu_C(\mu_C + a_r(L+1))$$
$$\sigma_x^2(L) \simeq \frac{a_r^2 L^2}{12} + \sigma_n^2(L) \tag{6}$$

where a_r and b_r are the amplitude and the intersect of the residual trend (i.e. remaining tectonic rate). Now, if we assume that the remaining seasonal signal is a pseudo periodic function at frequencies similar to the seasonal signal in Equation (1), hence taking the form $s_r(t) = \sum_{j=1}^{N} c_{r,j}\cos(d_j t) + e_{r,j}\sin(d_j t)$. Thus, we can do the same estimation as above in the case of a remaining pseudo periodic component in the residual time series, such as:

$$x(t_i) = s_r(t_i) + n(t_i)$$
$$< x(L) > = \frac{1}{L}\sum_{i=1}^{L}(s_r(t_i) + n(t_i))$$
$$< x(L) > = \delta + \mu_C$$

$$\sigma_x^2(L) = \frac{1}{L}\sum_{i=1}^{L}(x(t_i) - < x(L) >)^2$$
$$\sigma_x^2(L) = \frac{1}{L}\sum_{j=1}^{N} c_{r,j}^2\cos^2(d_j t) + e_{r,j}^2\sin^2(d_j t) +$$
$$\sigma_n^2(L) - \frac{2}{L}\sum_{i=1}^{L}(\sum_{j=1}^{N} c_{r,j}\cos(d_j t))(\sum_{j=1}^{N} e_{r,j}\sin(d_j t)))$$
$$- < x(L) >^2$$
$$\sigma_x^2(L) \simeq \sigma_n^2(L) + \sum_{j=1}^{N}\frac{c_{r,j}^2}{2} + \frac{e_{r,j}^2}{2} \tag{7}$$

$\simeq$ means that we assume $L >> 1$. From Equation (2), the mean of the noise is considered to be equal to the mean of the coloured noise μ_C. δ is the average of the remaining seasonal signal. It is assumed to be independent of L and bounded such as a periodic function.

We assume that the deterministic signals and the noise are completely uncorrelated, which is valid with white Gaussian noise (i.e. Wiener process) in signal processing (Papoulis and Unnikrishna Pillai , 2002; Montillet et al. , 2013). As previously discussed in Section 2.1, coloured noise can generate long- memory processes, hence producing non-zero covariance with residual signals. Due to the

varying amplitude of the coloured noise in geodetic time series of mixed spectra, the uncorrelated assumption is currently debated within the community (Herring et al. , 2016; He et al. , 2017). Therefore, recent works have introduced a random component together with a deterministic signal (i.e. nonlinear rate (Dmitrieva et al. , 2017), non deterministic seasonal signal (Davis et al. , 2012; Chen et al. , 2015; Klos et al. , 2018)). Thus, strictly speaking, $\sigma_x^2(L)$ should be seen as an upper bound.

The closed-form solution of the variance $\sigma_x^2(L)$ shows that the variance is unbounded in the case of a residual linear trend. To recall the discussion in Section 3.1, if this residual trend originates from various sources not well-described in the functional and stochastic model (i.e. undetected jumps, small amplitude random-walk component) of the geodetic time series, the amplitude of this trend should be rather small ($a < 1$ mm/yr) considering the length of GNSS time series available until now ($L < 30$ years). Unless this nonlinear residual trend has a large amplitude, a correction of the functional model must be done a posteriori.

To conclude, this property of the variance $\sigma_x^2(L)$ limits the characterization as a stable Levy process of the stochastic properties of the residual time series, hence restricting the application of the family of α-stable distributions.

4 Conclusions

We have discussed the assumptions that the underlying stochastic processes belong to the family of Levy processes in order to justify the use of the α-stable distributions in previous works, such as Montillet and Yu (2015). To do so, we model the residual GNSS time series as a sum of three r.v. The first two are defined from the stochastic model and assumptions on the noise properties of the geodetic time series, considered here as mixed spectra. The third r.v. is assumed to belong to the Levy processes. We then distinguish three cases.

In the case of a residual time series containing only short-term processes, the r.v. is a Gaussian Levy process, best fitted with an ARMA model. In the presence of long-term correlations and exhibiting self-similarity property, fractional Levy processes can be seen as an alternative model of using the fBm, assuming that the increments remain stationary and Gaussian. Also, due to the linear relationship between the Hurst parameter within the fBm model and the the fractional parameter of the FARIMA, it is likely that the FARIMA can fit the residual time series under specific conditions (i.e. amplitude of the noise components (Montillet and Yu , 2015)). Stable processes have infinite variance, which can mainly be satisfied in the case of heavy-tailed distributions in the application to geodetic time series. This condition is generally satisfied if there is an important misfit between the models (i.e. functional and stochastic) and the observations. However, a careful analysis of the time series should be done a prirori.

Bibliography

Benassi A., Cohen S., and Istas J. (2004) On roughness indices for fractional fields, Bernoulli, 10(2), 357?373.

Blewitt G, Lavallée D (2002) Effect of annual signals on geodetic velocity. J. Geophys. Res., 107. doi:10.1029/2001JB000570

Bock Y, Melgar D (2016) Physical applications of GPS geodesy: a review. Rep. Prog. Phys., 79 (10). doi:10.1088/0034-4885/79/10/106801.

Botai OJ, Combrinck L, Sivakumar V (2011) Interferences of α-stable distribution of the underlying noise components in geodetic data. South Afr. J. of Geo.. doi:10.2113/gssajg.114.3-4.541

Chen Q., Weigelt M., Sneeuw N., van Dam T. (2015) On Time-Variable Seasonal Signals: Comparison of SSA and Kalman Filtering Based Approach. In: Sneeuw N., Nov?k P., Crespi M., Sans? F. (eds) VIII Hotine-Marussi Symposium on Mathematical Geodesy. International Association of Geodesy Symposia, vol 142. Springer, Cham

Cont R., Tankov P. (2004) Financial modelling with jump processes, Chapman & Hall/CRC, ISBN 1-58488-413-4

Contreras-Reyes JE, Palma W (2013) Statistical analysis of autoregressive fractionally integrated moving average models in R. Comput. Stat., 28(5):2309-2331. doi: 10.1007/s00180-013-0408-7

Davis JL, Wernicke BP, Tamisiea ME (2012) On seasonal signals in geodetic time series. J. Geophys. Res., 117 (B01403). doi:10.1029/2011JB008690

Dmitrieva K., Segall P., Bradley A. M. (2017), Effects of linear trends on estimation of noise in GNSS position time-series, Geophys. J. Int., 208(1), 281-288, doi:10.1093/gji/ggw391.

Gazeaux, J. et al. (2013) Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment, J. geophys. Res., 118(5), 2397?2407

Granger CW, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. J. Time Ser. Anal., 1:15?29

Haykin S (2002) Adaptive Filter Theory. fourth edition, Prentice Hall Upper Saddle River, New Jersey

He X, Montillet JP, Fernandes R, Bos M, Yu K, Hua X, Jiang W (2017) Review of current GPS methodologies for producing accurate time series and their error sources. J. Geodyn., 106. doi:10.1016/j.jog.2017.01.004.

Herring TA, King RW, McClusky SC, Floyd M, Wang L, Murray M, Melbourne T, Santillan M, Szeliga W, Phillips D, Puskas C (2016) Plate Boundary Observatory and Related Networks: GPS Data Analysis Methods and Geodetic Products, Rev. Geophys., 54. doi:10.1002/2016RG000529.

Hurst HE, Black R, Sinaika YM (1965) Long Term-Storage: An Experimental Study. Constable, London

N. Kasdin (1995) Discrete simulation of colored noise and stochastic-processes and $1/f^{\alpha}$ power-law noise generation, Proc. IEEE, vol. 83 (1995).

Klos A., Bos M.S., Bogusz (2018), Detecting time-varying seasonal signal in GPS position time series with different noise levels, J. GPS Solut, 22 (21), doi:10.1007/s10291-017-0686-6

Koutrouvelis I. A. (1980) Regression-type estimation of the parameters of stable laws, J. Am. Statist. Assoc., 75:918-928

Langbein, J. (2008), Noise in GPS displacement measurements from Southern California and Southern Nevada, J. Geophys. Res., 113, B05405, doi:10.1029/2007JB005247

Langbein J. (2012) Estimating rate uncertainty with maximum likelihood: differences between power-law and flicker–random-walk models J. Geod., 86(9). doi:10.1007/s00190-012-0556-5

Li J, Miyashita K, Kato T, Miyazaki S (2000) GPS time series modeling by autoregressive moving average method: Application to the crustal deformation in central Japan. Earth Planet Space, 52:155-162

Li T.-H. (2013) Time Series with Mixed Spectra, CRC Press. ISBN 9781584881766.

Mandelbrodt B, Van Ness JW (1968) Fractional Brownian Motions, Fractional Noises and Applications. SIAM Rev., 10(4):422-437

Marquardt T. (2006) Fractional? Levy processes with an application to long memory moving average processes. Bernoulli, 12(6), 1099?1126.

Montillet J-P, Tregoning P, McClusky S, Yu K (2013) Extracting white noise statistics in GPS coordinate time series. IEEE Geosci. Remote Sens. Lett., 10(3):563-567. doi:10.1109/LGRS.2012.2213576

Montillet JP, Yu K(2015) Modeling geodetic processes with levy α-stable distribution and FARIMA. Math. Geosci., 47(6). doi:10.1007/s11004-014-9574-6

Nikias CL, Shao M (1995) Signal processing with Alpha-Stable Distributions and Applications. New York, Wiley edition.

Nolan JP (2009) Stable Distributions - Models for heavy tailed data. Book online: $http://www.math.ucla.edu/\ biskup/275b.1.13w/PDFs/Nolan.pdf$

Panas E (2001) Estimating fractal dimension using stable distributions and exploring long memory through ARFIMA models in Athens Stock Exchange. Appl. Fin. Econ., 11(4):395-402

Papoulis A., Unnikrishna Pillai S. (2002) Probability, Random Variables and Stochastic Processes, 4th Ed., McGraw-Hill Series in Electrical and Computer Engineering, the McGraw-Hill companies. ISBN: 0-07-366011-6.

Tregoning P, Watson C (2009) Atmospheric effects and spurious signals in GPS analyses. J. Geophys. Res., 114 (B09403). doi:10.1029/2009JB006344

Wang C, Liao M, Li X (2008) Ship Detection in SAR Image Based on the Alpha-Stable Distribution. Sensors, 8:4948-4960. doi: 10.3390/s8084948

Williams SDP (2003) The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. J. Geod., 76:483-494

Williams SDP, Bock Y, Fang P, Jamason P, Nikolaidis RM, Prawirodirdjo L, Miller M, Johnson DJ (2004) Error analysis of continuous GPS position time series. J. Geophys. Res., 109(B03412). doi:10.1029/2003JB002741

Wooldridge JM (2010) Econometric Analysis of Cross Section and Panel Data. First edition, MIT Press, ISBN (13):9780262232586

Daily reference evapotranspiration forecasting for oceanic climate using autoregressive Hilbertian process

Rousseau Tawegoum[1], Besnik Pumo[2], and Pierre Santagostini[2]

[1] Environmental Physics and Horticulture (EPHor) Research Unit, Agrocampus Ouest, Angers, France,
[2] IRHS, Agrocampus Ouest, INRA, Universit dAngers, SFR 4207 QuaSaV, 49045 Angers, France
rousseau.tawegoum@agrocampus-ouest.fr

Abstract. Predicting daily reference evapotranspiration is particularly important for short-term irrigation management in constrained horticultural nurseries. This paper investigates reference evapotranspiration daily forecasting based hourly observations viewed as functional data recorded into constant interval of hours. A model is carried out within the framework of autoregressive Hilbertian process. A package developed using R software served as simulation tools. Simulation results using Fourier basis show accurate forecasting in comparison with the seasonal autoregressive integrated moving-average parametric model, with respect to mean squares error.

Keywords: Time Series, Functional Data, Autoregressive Hilbertian Process, Sarima Model, Limited Data, Evapotranspiration

1 Introduction

Horticultural and agricultural systems require an efficient scheduling tool for production control, including irrigation, a major consumer of water in the nurserie or on the farm. The ability to accurate forecast reference evapotranspiration (ETo) could enhance the capability to manage irrigation systems by predicting crops water need-ing.

ETo forecasting can be investigated using direct or indirect method, depending on the approach used and data input. Direct methods in which current and historical data is used to forecast ETo either using time series or neural networks (ANNs or CNN). In the indirect method, weather variables needed to compute ETo are forecasted (of-ten by Numerical Weather Prediction (NWP) models) and then empirical (as Hagreaves-Simani [1] or analytical (as Penman-Monteith [2]) are used to compute ETo.

Monthly or weekly ETo forecast are helpful for planning middle to long-term irri-gation, while daily ETo forecast are more useful in short-term irrigation scheduling. Daily forecast is now a trend A few studies dealt with daily ETo

forecasting using time series technique [3-6]. Over the last decades, intelligent computational models have been developed as an alternative method for estimating reference evapotranspiration, such as artificial neu-ral network (ANN) technique. In this area several studies have used ANNs to estimate (forecast) ETo as function of climatic variable, or based past historical values of ETo, for example [7] for arid climate, [8] under climate of alava region, Spain. Kisi [9] developed two different feed-forward neural network models for estimation of daily ETo from climatic data. Landeras et al. [8] compared ANN and alternative evapotran-spiration equations with lower input requirements to PM56 under climate of Alava region, Spain. Yassin et al. [7] investigated the application of ANN for modelling ETo for aride climate in Arabia peninsula.

Sometime after, Evolutionary algorithms such genetic programming or gene ex-pression programming (GEP) [10] a natural development of genetic algorithm and genetic programming have been used in reference evapotranspiration forecasting. Gene expression programming have been widely used for daily ETo forecasting, [7;11-12] evaluated the generalization GEP for estimating ETo in a coastal environ-ment for eight stations by comparison of locally and externally trained models , and concluded that the accuracy differences between the local and the external perform-ance depend on the specific climatic trends of the test stations. Yassin et al. [7] investigated the application of GEP and ANN for modeling daily ETo for arid climate showed that ANN technique were slightly more accurate than those developed using the GEP technique. Shiri [12] compared the GEP technique for estimating daily ETo values in an hyper-arid region of Iran, with the corresponding empirical or assessed the performance of different ETo models, which outperform empirical or semi-empirical model (for temperature/humidity radiation and combination based ap-proach).

By the beginning of the 21th century, the forecasting performance of mesoscale numerical weather prediction (NWP) model had enhanced consequently what induced the use of indirect method to forecast daily ETo. Some studies have attempted to fore-cast ETo using NWP or global climate model for a lead time up to five days [4,6,13-16]. The results showed that the forecast error increased as the number of forecasting days increased.

Similarly, studies have also been conducted for daily ETo forecasting from public weather forecast [17-19], in combination with ANN technique [20;21] or of GEP [19]. More precisely, [20] combined ANN and data from NWP up to three days to predict crop cater requirement of maize and onions for seven agrometeorological stations of Easter Mancha, Spain.

Most of these approaches are relied on daily mean data climatic variables observations in combination with PM or HS models, or relied on coarse-scale NWP model or global climatic model weather forecasts, less accurate on local scale. Availability of local hourly ETo observations could lead to accurate daily ETo computation. These approaches also required a fairly long record of data and observations for training and validation.

Indeed, computation of daily ETo from successive hourly observations from near-by meteorological stations could induce accurate results than those based

on mean daily observed climatic data. Thus, considering ETo as functional data might improve daily prediction. Functional data analysis extends the classical multivariate statistic to the case of functional data.

The purpose of this work is to propose a prediction method based on Functional Data Analysis (FDA) which has known a rapid development in the late 20th century. The book of Ramsey and Silverman [22] is the indispensable resource giving a large panorama on analysis of such data while the book of Bosq [23] is the reference for prediction methods based on FDA. Bosq in this book presents the autoregressive process taking values on some function spaces. The particular case of autoregressive processes taking values on some Hilbert space, denoted ARH, has been object of study for many researchers. The pioneering work on ARH processes is due to Bosq [24] which introduces these processes and give theoretical results concerning the es-timation and convergence for the unknown parameter and unobserved prediction. Alternative approaches, extensions and numerical implementations with theoretical necessary complements was proposed by [25-34]. We invite the reader to consider the bibliography cited in the paper [35] for a more complete list of work on ARH proc-esses and [36] for some more recent developments.

The paper is organized as follows: the first part defines ETo, the second introduces forecasting models used, and the last presents simulation results.

2 Reference evapotranspiration modeling

Evapotranspiration (ET) is defined as the evaporation from a soil surface and the transpiration from plant material [2]. ET is part of a balanced energy budget that ex-changes energy for outgoing water at the plant surface. Parameters that drive ET are solar radiation, temperature, relative humidity and wind speed.

Reference ET (ET0) is defined as the ET from a hypothetical reference crop with the characteristics of an actively growing, well-watered, dense green cool season grass of uniform height. There exist many equations for estimating ET0, but the inter-national scientific community has accepted the Penman-Monteith (P-M) equation as the most precise one for its good results compared with other equations in various regions worldwide [2].

The FAO-56 PM equation for the hourly time step reads as follows:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma(\frac{37}{T_{hr}+273})U_2(e^0(T_{hr}) - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \tag{1}$$

where ET_0 is the reference evapotranspiration ($mm\,hr^{-1}$); Δ the slope of the satura-tion vapor pressure curve at T_{hr} ($kPa\,^\circ C^{-1}$); R_n net radiation at the grass surface ($MJ\,m^{-2}\,h^{-1}$); G the soil heat flux density ($MJm^{-2}h^{-1}$); γ the psychrometric constant ($kPa\,^\circ C^{-1}$); T_{hr} the mean hourly air temperature ($^\circ C$); U_2 the average hourly wind speed ($m\,s^{-1}$); $e^0(T_{hr})$ the saturation vapor pressure at T_{hr} (kPa); e_a is the actual average hourly vapor pressure (kPa).

For daily time step, the '37' value in Eq. (1) changes to '900' for ETo in ($mm\,day^{-1}$), Rn and G ($MJ\,m^{-2}\,day^{-1}$), $e^0(T_{hr})$ is computed as $(e^0(T_{max}) + e^0(T_{min}))/2$, where e^0 is the saturation function and T_{max} and T_{min} are daily maximum and minimum air temperature.

3 Time series modeling

3.1 Sarima time series

The statistical prediction of a time series is used in the industrial, climatic, medical and environmental fields, and in open and furrow irrigation systems [37-39]. We used the estimation procedure suggested by Box et al., [40] in this context.

The general multiplicative SARIMA model of order $(p, d, q) \times (P, D, Q)_S$ is defined as follows:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D z_t = \theta_q(B)\Theta_Q(B^S)a_t \tag{2}$$

where

$$\nabla = 1 - B \tag{3}$$

$$\nabla_S = 1 - B^S \tag{4}$$

- B is the backward shift operator,
- z_t are formed from the sampled original series at time t,
- a_t is a purely random process (corresponding to a zero-mean Gaussian white noise with variance σ_e^2)
-∇ is the differencing operator, S is the period of the series, d is the ordinary differencing order and D is the seasonal differencing order.
∇_S is the seasonal differencing
ϕ_p, Φ_P, θ_q and Θ_Q are polynomials in B of order p, P, q and Q, respectively, that fulfill the stationarity and invertibility condition.

P and Q are the degrees of the autoregressive and moving average seasonal polynomials , p and q are the degrees of the autoregressive and moving average polynomials . The selected model is expected to forecast the next value of ET0, based on past measurements. The forecast model was obtained by the iterative strategy of specification, estimation and checking. The values of p, P, q and Q were thus assessed by studying the autocorrelation function (acf) and the partial autocorrelation function (pacf) of the differenced series, and by choosing a SARIMA model where acf and partial acf had a similar form [40]. d, D, p, q, P and Q values were computed using R software package.

3.2 Autoregressive Hilbertian process modeling

The model generalizes the classical AR(1) process for scalar or multivariate time series to functional data. It was introduced by [24]. Let $X_1, \cdots, X_n$ be a sample

of random curves for which a stochastic dependence is suspected (for instance the curve of temperature observed during n days at a given place). We shall assume that all the X_i-s belongs to some separable Hilbert space H endowed with scalar product $\langle \cdot, \cdot \rangle$ and norm $|| \cdot$
. An autoregressive process of order one, denoted $ARH(1)$, is a H-valued process $(X_i, i \in Z)$ satisfying

$$X_i = \rho(X_{i-1}) + \epsilon_i \tag{5}$$

where

- ρ is a linear operator from H to H such that $||\rho_{j0}||$ for some positive integer j_0,
- $(\epsilon_i)_{i \in Z}$ is a sequence of independent and identically distributed H-valued variables with common covariance operator (see [32] for instance for the definition of covariance operator in Hilbert spaces) and such that $E(\epsilon_i | X_{i-j}, j > 1) = 0$.

The model is simple, with a single unknown operator, leaving however the possibility to decline various assumptions either on the operator ρ. Bosq [23] shows that 5 admits a unique stationary solution, which is a Markov process. As often noted, the interest of the model relies in its predictive power. The estimation of ρ is the first and necessary step before deriving the statistical predictor of X_{n+1}. The estimation of ρ is based on some Yule Walker equivalent equations for real AR process (see [42]). So, let us first define the covariance operator of X_1 as an operator defined on H by $\Gamma(\cdot) = E[\langle X_1, \cdot > X_1 \rangle]$ where E is the expectation operator defined on H. The crossed covariance operator of X_2, X_1 is given by $\Delta(\cdot) = E[\langle X_2, \cdot > X_1 \rangle]$ defined on H. On can show that $\Gamma\rho = \Delta$. The estimated operators of Γ and Δ are respectively

$$\Gamma_n(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \langle X_i, \cdot \rangle X_i \quad \text{and} \quad \Delta_n(\cdot) = \frac{1}{n-1} \sum_{i=1}^{n-1} \langle X_{i+1}, \cdot \rangle X_i \tag{6}$$

The real interest for ARH processes is when the dimension of H is infinite, since otherwise the space is isomorph to Euclidean space R^d, d being the dimension of the space. In the infinite case the difficulty to estimate ρ comes from the fact that the Yule Walker equation is an ill posed problem. Bosq tackle this difficulty by projecting data in some finite Hilbert subspace and suitable choosing the dimension k of this sub-space. Let us briefly define the estimator of ρ which will be denoted $\hat{\rho}_{n,k}$ in the following and the corresponding predictor $\hat{X}_{n+1,k}$. Let $(\phi_{n,j}, j \geq 1)$ be a complete orthonormal system of eigen-functions of Γ_n and $(\lambda_j, j \geq 1)$ in decreasing order the corresponding eigen-values. Denoting V_k the subspace of H generated by the first eigen-functions $\phi_{n,j}$ and Π_k the orthogonal projection operator on V_k the estimator of ρ is (see for exemple Bosq [24]) :

$$\hat{\rho}_{n,k} = (\Pi_k \Gamma_n \Pi_k)^{-1} (\Pi_k \Delta_n \Pi_k) \tag{7}$$

It depends on n and k the dimension of the subspace V_k.

Bosq proves under mild conditions that the predictor converges to when n and k go to infinity (note that k depends on n and sometimes one notes $k_n$in place of k).

Alternative approaches are proposed by [25], [41] for the estimation of ρ based on an idea that consists to add a small perturbation in order to regularize the un invertible covariance operator (see Tikhonov and Arsenin, [43]).

4 Simulation results

In order to compute hourly ET0, a climatic database with four types of measurements (global radiation, air temperature, relative humidity and wind speed) was used. These measurements were made in the INRA agrometeorological station of Montreuil-Bellay (47 07 56 N. 00 09 08 W), neighboring the city of Angers in the West region of FRANCE. The climate in this area is typically oceanic (cool and relatively humid summers), with continental influences (wide temperature ranges). The hardest period for the crops (corresponding to maximum water loss) extended from June to September, 2015. This period included steady-state weather, but also cloudy skies and unsettled weather with brief spells of rainfall. The first part of data was used for identification and the second for validation forecasting.

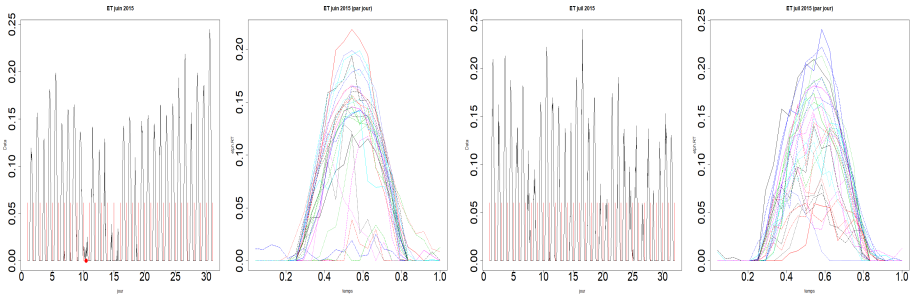


Fig. 1. Identification and validation data

From a technical point of view the different approach for implementing an ARH proceed in two steps. The first step consists in decomposing data in some functional basis in order to reconstruct them on the whole observed interval. Most of the methods [26,34,25] use spline or wavelet basis and suppose that the curves belong to the $W^{2,k}$ space of functions such that the k-th derivative is squared integrable. The second step consists in choosing tuning parameters required by these methods, for exam-ple the dimension of the projection subspace for the projection estimations. A general method used is cross-validation.

A software has been carried out using R environment [44]. The model structure for sarima was of order $(1,0,0) \times (1,1,0)_{24}$ (see $eq.2$), and for ARH with Fourier basis, $k = 1$ (see $eq.7$).

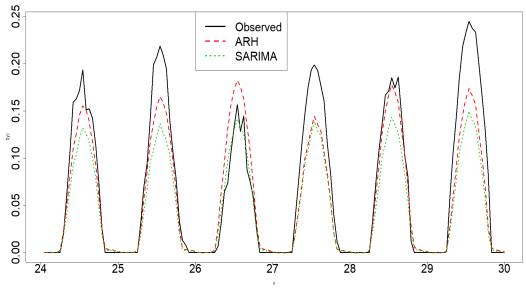


Fig. 2. validation: ETo forecasting June 25-31, 2015

Data from June 1-23 was used for identification, and those for June 24-30 for vali-dation. For this period obtained models under predict daily ETo except on day 27 where ARH(1) forecast is above observation. However, in most of the days, ARH gives better results compared to sarima model.

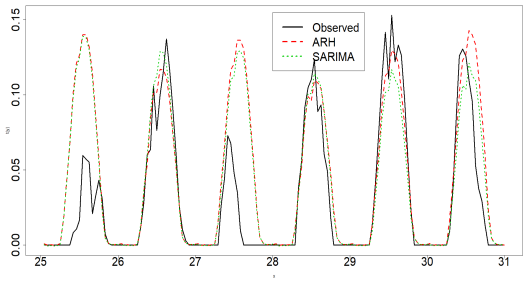


Fig. 3. validation: ETo forecasting July 26-31, 2015

Figure 3 shows forecasts validation for days of July 26-31, with training data of July 1-24. Daily observation has different forms each other, due to unsettled weather conditions. For example, observed value of days 26 and 28 have particular low mag-nitude compared to standard values expected for this period. In this context, model fails to capture in advance these behaviors. ARH gives better results for days 29 and 30, while sarima model is slightly better for day 28.

Figure 4. presents forecasts validation results for July 26-31 with training data ex-tending from June 1rt to July 25. ARH(1) gives results similar to those of the previous case. For sarima model with long training data validation results are slightly better than those with short training data.

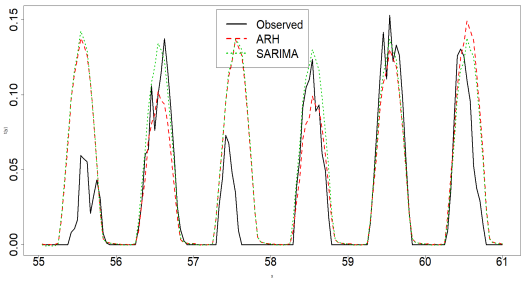


Fig. 4. validation:ETo forecasting July 26-31, 2015.(long identification data)

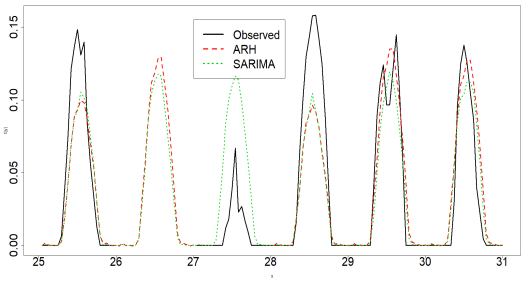


Fig. 5. validation: ETo forecasting August 26-31, 2015

Figure 5 presents validation results for August 2015. One can observe that due to missing observation of day 27, prediction errors or daily ETo forecasts for day 28 can not be evaluated. ETo forecast for days 30 and 31 with ARH(1) are much close to observed values in comparison with sarima model. The observed and

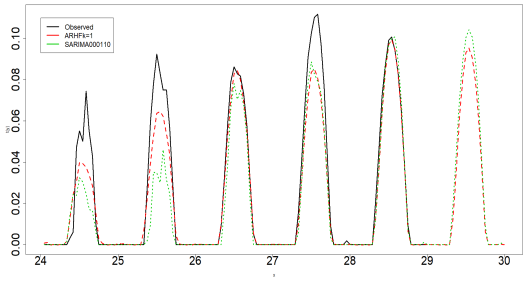


Fig. 6. validation: ETo forecasting Sept 25-29, 2015

forecast daily values of ETo were compared using mean-squared error criteria, provided in table 1.

Table 1. validation forecast mean squared errors

	day	N–5	N–4	N–3	N–2	N–1	N	mean
June, N=30	SARIMA	0.009	0.017	0.017	0.026	0.003	0.052	0.021
	ARH	0.008	0.016	0.017	0.027	0.002	0.052	0.020
July, N=31	SARIMA	0.071	0.005	0.081	0.003	0.004	0.023	0.031
	ARH	0.071	0.003	0.076	0.004	0.003	0.022	0.030
Aug, N=31	SARIMA	0.008	NA	NA	0.052	0.014	0.011	0.021
	ARH	0.013	NA	NA	0.024	0.007	0.010	0.014
Sept, N=30	SARIMA	0.007	0.015	0.001	0.004	0.000	NA	0.005
	ARH	0.003	0.003	0.000	0.004	0.000	NA	0.002

5 Conclusion

We have proposed in this study, daily ET0 forecasting based on functional data analysis. The study was conducted with data from local near-by agro-meteorological station, now a day, easily accessible. Compared to standard men-

tioned techniques, the approach does not need a long historical data. The approach used produces an accu-rate forecasting in steady-state, but the performance decreases somewhat during un-settled weather conditions. Future work could consist to get better results with ARH by differentiation of raw data before estimation. The behaviour of the model on the simulated data encourages the use of the approach for short-term irrigation scheduling.

References

1. Hargreaves GH, Samani ZA (1985) Reference crop evapotranspiration from temperature. Appl. Eng. Agric. 1 (2), 96-99
2. Allen R.G, Pereira L.S, Raes D, Smith M (1998) Crop evapotranspiration: Guidelines for computing crop requirements. Irrigation and Drainage Paper N 56 FAO, Rome Italy, 300pp.
3. Hess T (1996) A microcontroller scheduling program for supplementary irrigation. Computers and Electronics in agriculture. (15) 233-243.
4. Duce P, Snyder RL, Spano D (2000) Forecasting reference evaporation. Proceed-ings of 3rd International Symposium on Irrigation horticultural crops. Int. Soc. Hortic. Sci. ISHS 537 (1), 135-141.
5. Trajkovic S, Todorovic B, Stankovic M (2003) Forecasting of reference evapotranspiration by artificial neural network, J. Irrigation and drainage engineering 129 (6) 454-457
6. Arca B, Duce P, Spano D, Snyder RL, Fiori M (2004) Use of numerical weather forecast and time series models for predicting reference evapotranspiration In Proc IVth Int Symp Irrigation of Horticultural Crops Ed Snyder RL, Acta Horticulturae 664, 39-46.
7. Yassin MA, Alazba AA, Mattar MA (2016) Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid cli-mate Agricultural water management 163, 110-124
8. Landeras G, Ortiz-Barredo A, Lopez JJ (2008) Comparison of artificial neural networh models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque country (Northern Spain) Agricultural water management 95, 553-565
9. Kisi O (2016) Modeling reference evapotranspiration using three different heuristic regression approaches, Agricultural water management 169, 162-172
10. Ferrera C. (2011) Gene expression programming:a new algorithm for solving problems. Complex Syst. 13 (2) 87-129
11. Shiri J, Sadraddini AA, Nazemi AH, Kisi O, Landeras G, Fard AF, Marti Pau (2014) Generalizability of gene expression programming-based approaches for estimating daily evapotranspiration in coastal station of Iran J. Hydrology 508, 1-11
12. Shiri J (2017) Evaluation of FAO-PM empirical semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid region of Iran Agricultural water management 188, 101-114
13. Perera KC, Wester AW, Nawarathna B, George B (2014) Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs Agricultural water management 194, 50-63
14. Tian D, Martinez CJ (2012) Comparison of two analog-based downscaling methods for regional reference evapotranspiration forecasts. J. Hydrol. 475 350-364

15. Ishak AM, Bray M, Remesanm R, Han D (2010) Estimating reference evapotranspiration using numerical weather modelling. Hydrol. Process. 24 (24) 3490-3509
16. Silva D, Meza FJ, Varas E (2010) Estimating reference evapotranspiration using numerical weather forecast data in centre Chile J Hydrology 382, 64-71
17. Cai J, Liu Y, Lei T, Pereira LS (2007) Estimating reference evapotranspiration with the FAO Penman-Monteith equation using daily weather forecast messages. Agri. Forest Meteorol. 145 (1-2) 22-35
18. Luo Y, Chang X, Peng S, Hkan S, Wang W, Zeng Q, Cai X (2014) Shirt-term forecasting of daily reference evapotranspiration using the Hagreaves-Samani model and temperature forecasts. Agricultural water management 136, 42-51
19. Yang Y, Cui Y, Luo Y, Lyu X, Traor S, Khan S, Wang W (2016) Forecsting reference evapotraspiration using Penman-Monteith model and public weather forecasts. Agricultural water management 177, 329-339
20. Ballesteros R, Ortega JF, Mareno NA (2016) Foreto new software for reference evapotranspiration forecasting, J. arid environments, 124, 128-141
21. Traor S, Luo Y, Fipps G (2016) Deployment of artificial neural network for short-term forecasting of evapotranspiration using public weather forecast restricted messages Agricultural water management 163, 363-379
22. Ramsay J.O., Silverman B.W. (1997). Functional Data Analysis, Springer Series in Statistics.
23. Bosq D. (2000). Linear processes in function spaces. Lectures notes in statistics, Springer Verlag.
24. Bosq D. (1991). Modelization, nonparametric estimation and prediction for continous time processes. in Roussas (Ed), Nato Asi Series C, 335, 509-529.
25. Antoniadis A. and Sapatinas T. (2003). Wavelet methods for continuous-time predction using representations of autoregressive processes in Hilbert spaces. J. of Multivariate Analysis, 87, 133-158.
26. Besse P. and Cardot H. (1996). Approximation spline de la prvision d'un pro-cessus fonctionnel autorgressif d'ordre 1, Canad. J. Statist, 24, 467-487.
27. Besse P., Cardot, H. and Stephenson, D. (2000). Autoregressive forecasting of some climatic variations. Scand. J. Statist, 27, 673-687.
28. Kargin V. and Onatski A. (2008). Curve forecasting by functional autoregression. J. of Multivariate Analysis, 99, 2508-2526.
29. Damon J., Guillas, S. (2005). Estimation and Simulation of Autoregressive Hilbertian Processes with Exogenous Variables. Stat. Infer. for Stoch. Proc., 8(2), 185-204.
30. Labbas A. and Mourid, T. (2002). Estimation and prediction of a Banach valued autoregressive process. C. R. Acad. Sci. Paris, 335(9), 767-772.
31. Marion J.M., Pumo B. (2004). Comparaison des modles ARH(1) et ARHD(1) sur des donnes physiologiques, Annales de l'ISUP, 48(3), 29-38.
32. Mas A. and Pumo, B. (2007). The ARHD process. J. of Statistical Planning and Inference, 137, 538-553.
33. Merlevde, F., Peligrad M., Utek, S. (1997). Sharp conditions for the CLT of linear processes in Hilbert space. J. of Theoritical Probability, 10(3) 681-693.
34. Pumo B. (1998). Prediction of continuous time processes by C[0,1]-valued autoregressive process. Statist. Infer. for Stoch. Processes, 3(1) 297-309.
35. Mas A. and Pumo, B. (2010) Linear processes for functional data, in The Oxford Handbook of Functional Data Analysis, edited by Frdric Ferraty and Yves Ro-main
36. Horvath L., Huskova M and Kokoszka P. (2010). Testing the stability of the functional autoregressive process. J. of Multivariate Analysis, 101, 352-367.

37. Brockwell P. J, Davis R.A (1996) Introduction to time series and forecasting. Springer, 420p.
38. Crosbie R.S, Binning P. and Kalma J.D (2005) A time series approach to inferring ground water recharge using the water table fluctuation method. Water Resources Research 41, W01008, doi 1029/2004WR003077.
39. Young P.C (2006) New approaches to volcanic time series analysis. In Mader H.M, Coles S.G., Connor C.B. and Connor L.J. (Eds.), Statistic in volcanology, Bath, UK, pp. 143-160, Geological Society of London.
40. Box G. E P, Jenkins G. M, Reinsel G. C (1994) Time series analysis, forecasting and control. 3rd Edn., Englewood Cliffs, NJ: prentice-Hall.
41. Mas A., Menneteau L. (2003). Perturbation appraoch applied to the asymptotic study of random operators. Progress in Probability, 55 127-134.
42. Brockwell P. and Davis A. (1991). Time series: Theory and methods. Springer Verlag.
43. Tikhonov A.N. and Arsenin V.Y. (1977) Solutions of ill-posed problems, V.H. Winstons and sons, Washington.
44. R: A language and environment for statistical computing

Characterization and detection of potential fraud taxpayers in Personal Income Tax using data mining techniques.

Camino González Vasco
Dirección de Estudios
Instituto de Estudios Fiscales –Universidad Rey Juan Carlos
Avenida Cardenal Herrera Oria 378, 28035 Madrid
Email: camino.gonzalez@ief.minhafp.es

Mª Jesús Delgado Rodríguez
Departamento de Economía Aplicada II
Universidad Rey Juan Carlos
Campus de Vicálvaro, 28032 Madrid
Email: mariajesus.delgado@urjc.es

Sonia de Lucas Santos
Departamento de Economía Aplicada (Udi de estadística)
U. Autónoma de Madrid
Campus de Cantoblanco, 28049 – Madrid
Email: sonia.delucas@uam.es

(preliminary version)

May, 2018

Keywords: Tax evasion; Data mining techniques; Multilayer Perceptron; Decision Trees; Fiscal fraud detection; Tax evaluation;

JEL Classification: H26 Tax Evasion and Avoidance; C55 Large Data Sets: Modeling and Analysis; C38 Classification Methods; Cluster Analysis; Principal Components; Factor Models

ABSTRACT:

This paper contributes to the empirical literature on the development and estimation of a classification model for Personal Income Tax (PIT) evasion detection focusing on the study of the deductible expenses and reductions in the tax base that lead to this behavior. In this regard, the purpose of this study is twofold. Firstly, it attempts to determine those deductible expenses and reductions in the tax base that are more likely to be used by potential fraud taxpayers by means of investigating the design of PIT structure. Secondly, it aims at characterizing through socioeconomic variables the profile of the potential fraud taxpayer to provide an audit selection strategy for improving tax compliance. We apply a two-step methodology combining dimension reduction and classification data mining techniques to obtain a sample segmentation according to potential fraud probability. Using the micro data contained in the annual Spanish PIT sample designed by the Institute for Fiscal Studies, we identify hidden patterns in data and select suspicious evasion reports on deductible expenses and reductions in the tax base for further auditing and design considerations. The conclusion provides evidence that the combination of these data mining techniques is a promising alternative to potential fraud prediction and tax design evaluation.

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Spanish Institute for Fiscal Studies, the Department of Applied Economics II at the Rey Juan Carlos University of Madrid and the Department of Applied Economics (Statistics Unit) of the Autonomous University of Madrid.

0. Introduction.

Fiscal fraud detection is an important issue that incurs expenses in terms of the loss of government revenues, which leads to less efficient tax programs and the inequity between evaders and honest filers. Effective control of income tax evasion requires addressing a fundamental statistical problem of non detection, which can bias estimates of the overall amount of evasion and the relative evasion propensities of different socioeconomic groups.

Traditional ways of data analysis have been in use for a long time as a method of detecting fiscal fraud and revenue potential. Both supervised learning methods (where a dependent variable is available for training the model) and unsupervised learning methods (where no prior information of dependent variable is available for use) can be potentially employed to solve this problem.

In this paper we focus on exploring the potential applicability of the data mining techniques in developing a segmentation model that can detect and predict fraud suspicion in personal income tax micro data from deductible expenses and reductions information. Unlike previous studies focused in donations and tax benefits, deductible expenses and reductions of the tax base are are particularly interesting tax items for taxpayers to reduce the payment of this tax. The model is based on invidual level data drawn from the annual Spanish PIT sample designed by the Institute for Fiscal Studies.

Data mining techniques include different methodologies aimed at exploring and analyzing large quantities of data in order to discover meaningful patterns and rules, oriented to classification and prediction. Despite its importance, most of the data mining models developed for the detection of fraud are focused on financial fraud detection, and hence there is a lack of applications published in official tax data aimed at fiscal fraud detection.

That being the case, data mining techniques have been applied most extensively to the detection of insurance, corporate and credit card fraud but mortgage fraud, money laundering, and securities and commodities fraud have also attracted a great deal of attention in recent years. The main data mining techniques used for financial fraud detection are logistic models, artificial neural networks, the Bayesian belief network, and decision trees, all of which provide primary solutions to the problems inherent in the detection and classification of fraudulent data.

The classification model presented in this paper performs, as a prior step, a principal component analysis (PCA) as a dimension reduction technique of the group of economic and demographic variables to obtain a group of orthogonal indicators. Reducing or eliminating statistical redundancy between the components of high-dimensional vectors enables a lower-dimensional representation without significant loss of information. Secondly, we test Multilayer Perceptron (MLP) to assign

scores /probabilities that result in fiscal fraud. Finally we use decision trees (CHAID) to segment the Spanish PIT sample records according to the probabilities assigned by the MLP.

The purpose of this study is twofold. Firstly, it attempts to determine those deductible expenses and reductions in the tax base that are more likely to be used by potential fraud taxpayers. In this context this paper uses Principal Component Analysis to investigate the design of tax structure and will be of interest for making proposals aimed at improving the design of the Personal Income Tax Using the orthogonal factors resulting from the Principal Components Analysis, the Multilayer Perceptron scores more than two million PIT records corresponding to 2013. These Neural Network scores can be of use as an early warning indicator for potentially fraudulent activity. The CHAID segmentation of the taxpayers universe according to these scores is the last part of the methodology and allows for the second main objective of the study: characterizing through socioeconomic variables the profile of the potential fraud taxpayer to provide an audit selection strategy for improving tax compliance.

This paper is divided in six sections and proceeds as follows. Section I specifies the literature review. Section II presents how data mining techniques have facilitated the detection of tax evasion in tax administration. Section III describes the Spanish PIT Microdata included in the analysis. Section IV illustrates the algorithm used to generate the target variable of the model and points out the limitations of the study. Section V outlines the estimation technique and the empirical results obtained. Section VI evaluates the overall performance of the model and presents the segmentation results. Section VII concludes.

I. Literature Review.

In recent years, the techniques of data mining and artificial intelligence have been incorporated into the audit planning activities mainly to detect patterns of fraud or evasion, which are used by tax authorities for specific purposes. Most of the papers mentioned focus on Corporate Income Tax and Value Added Tax but there are few applications designed for Personal Income Tax.

Examples include examining how commodity flows respond to destination sales taxes, allowing for tax evasion as a function of distance between trade partners (Fox et al. 2014), identifying clusters or groups of taxpayers who have similar behavior using a clustering algorithm named the SOM method (González et al. 2013), using association rules to develop a screening model, mining tax evasion patterns from the universe of taxpayers (Wu et al. 2012), applying data mining techniques such as Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic programming (GP) , Probabilistic Neural Networks (PNN), and Logistic Regression (LR) to identify companies that resort financial statement fraud (Ravisankar et al.2011), testing a list of input variables and compare the performance of techniques as Classification trees, Logistic Regression and Discriminant Functions to assign records to fraud /non-fraud cases (Gupta, et al. 2007), using Linear Regression and Neural Network models for designing an audit selection strategy for the State of Texas (Micci-Barreca et al.2004), using Backpropagation Neural Networks, Recurrent Neural Networks and Artificial Immune Systems for fraud detection (Weatherford et al. 2002), taking advantage of clustering and classification techniques for constructing profiles of fraudulent behaviour, aimed at supporting audit planning (Bonchi et al.1999), among others.

Despite these screening and classification models for detecting fraud patterns oriented at audit planning, there are few studies that focus on the identification of those tax items or tax amounts in the PIT structure that are more likely to be used by potential fraud taxpayers .This study aims to fill this gap by developing a data mining framework based on a combination of Principal Components Analysis, Neural Networks and Decision Trees to identify most influential tax items for fraud detection and generate useful rules for potential fraud characterization.

II. Data mining methods for fiscal fraud detection.

Micci-Barrera et al. (2004) list and classify data mining applications in tax administration as follows: *Audit selection* is one of many possible data mining applications in tax administration. In the area of tax compliance alone, tax collectors face diverse challenges, from underreporting to nonreporting to enforcement. Data mining offers many valuable techniques for increasing the efficiency and success rate of tax collection. Data mining also leverages the variety and quantity of internal and external data sources available in modern data warehousing systems. Here are just a few of the many potential uses of data mining in the context of tax compliance:

Outlier-based selection: This is an alternative approach to audit selection that uses specific data mining algorithms to build profiles of typical behaviors and then select taxpayers that do not match the profiles. This modeling process involves creating valid taxpayer segments, characterizing the normative taxpayer profile for each segment, and creating the rules for outlier detection. Though this method is similar to that applied by many audit selectors on a daily basis, the added benefit of the data mining approach is its ability to process large amounts of data and analyze multiple taxpayer characteristics simultaneously.

Lead prioritization: Many tax agencies flag potential nonfilers by, for example, cross-matching external data with internal lists of current filers. This type of application often produces many leads, each of which must be manually verified and pursued. Given the limited staff available to follow up on leads, it is important to develop criteria for prioritizing the leads. Data mining, however, can predict the tax dollars owed by an organization or individual taxpayer by modeling the relationship between the attributes and reported tax for known taxpayers.

Profiling for cross-tax affinity: Agencies can use data mining to analyze existing taxpayers for associations between the types of businesses and the tax types (more than 30 in Texas) for which the taxpayers file. Co-occurrence of certain tax types may infer liability for another tax type for which the taxpayer did not file.

Workflow analysis : Following up on leads can be a lengthy process consisting of multiple mail exchanges between the tax agency and the organizations classified as potential nonfilers. Tracking this process, however, generates data that is useful for modeling the process. The process models can then be used to predict the value of current pipelines and to optimize resource allocation.

Anomaly detection : Taxpayers self-report several important attributes (SIC, organization type, etc.). Due to unavoidable data entry errors, however, some taxpayers are categorized incorrectly or even

uncategorized. Since these attributes and categories drive audit selection and other processes, it is always a good idea to apply data mining's rule induction techniques to detect errors and anomalies.

Economic models for optimal targeting: In most traditional targeting applications, such as target marketing or e-commerce fraud detection, there is a tradeoff between the number of audits to target and the cost per audit. When audit cost information is available, agencies can use data mining to develop targeting strategies that maximize overall collections.

III. The data set.

Our starting point is the microdata contained in the annual Spanish PIT sample. In this particular paper we use the sample for the year 2013, which includes 2.161.647 records extracted from a population of 19.203.031 registers providing personal income tax returns (Picos Sánchez, F. 2014). This database has been developed by the Spanish Institute of Fiscal Studies (Instituto de Estudios Fiscales, IEF henceforth), in collaboration with the Spanish National Tax Administration (Agencia Estatal de Administración Tributaria - henceforth AEAT), the entity in charge of extracting annual samples from its administrative registers of Spanish personal income tax.

For the construction of this annual sample the minimum variance stratification under Neyman's allocation method has been used. Three stratification variables have been used in the sampling process:

a) Territorial stratum: the province. 46 provinces with commun fiscal regime plus the Autonomous Cities of Ceuta and Melilla in addition to Non-resident taxpayers who are taxed under Article 10 of Law 35/2006.

b) Income stratum: income level of the tax filers. The sample income was calculated as the sum of net incomes, imputed income and capital gains and losses.

The sample income was divided into 12 groups for stratification:

- *Income group 1*: negative up to 0 euros.
- *Income group 2*: up to 6.000 euros.
- *Income group 3*: from 6.000,01 to 12.000 euros
- *Income group4*: from 12.000,01 to 18.000 euros
- *Income group5*; from 18.000,01 to 24.000 euros
- *Income group6*: from 24.000,01 to 30.000 euros
- *Income group7*: from 30.000,01 to 36.000 euros
- *Income group8*: from 36.000,01to 42.000 euros
- *Income group9*:from 42.000,01 to 48.000 euros
- *Income group10*: from 48.000,01 to 54.000 euros
- *Income group11*: from 54.000,01 to 60.000 euros
- *Income group12*: more than 60.000 euros.

c) The type of tax return stratum: separate or joint.

Hence, the "original weight" is calculated for each observation as the ratio between the

size of the population of its belonging stratum *h* and its corresponding sample size , $w^h = N^h / n^h$. To select the sample, the tax returns were classified in each of the 49x12x2=1.176 strata. The sample size *n* was calculated for a sampling error (in the average of the income variable) less than 1,1% with a confidence level of 3 per 1,000. A restriction of statistical confidentiality has been imposed on design sizes. Therefore, the population for each stratum (*Nh*) was determined using the population quasi-variance of the sample income S_h^2. Finally, using the values *Nh*and S_h^2, the sample size of each stratum n_h was determined so that $\sum_h n_h = n$.

Table1: Final micro-data sample sizes and their distribution by income (sum of net incomes, imputed income and capital gains and losses)

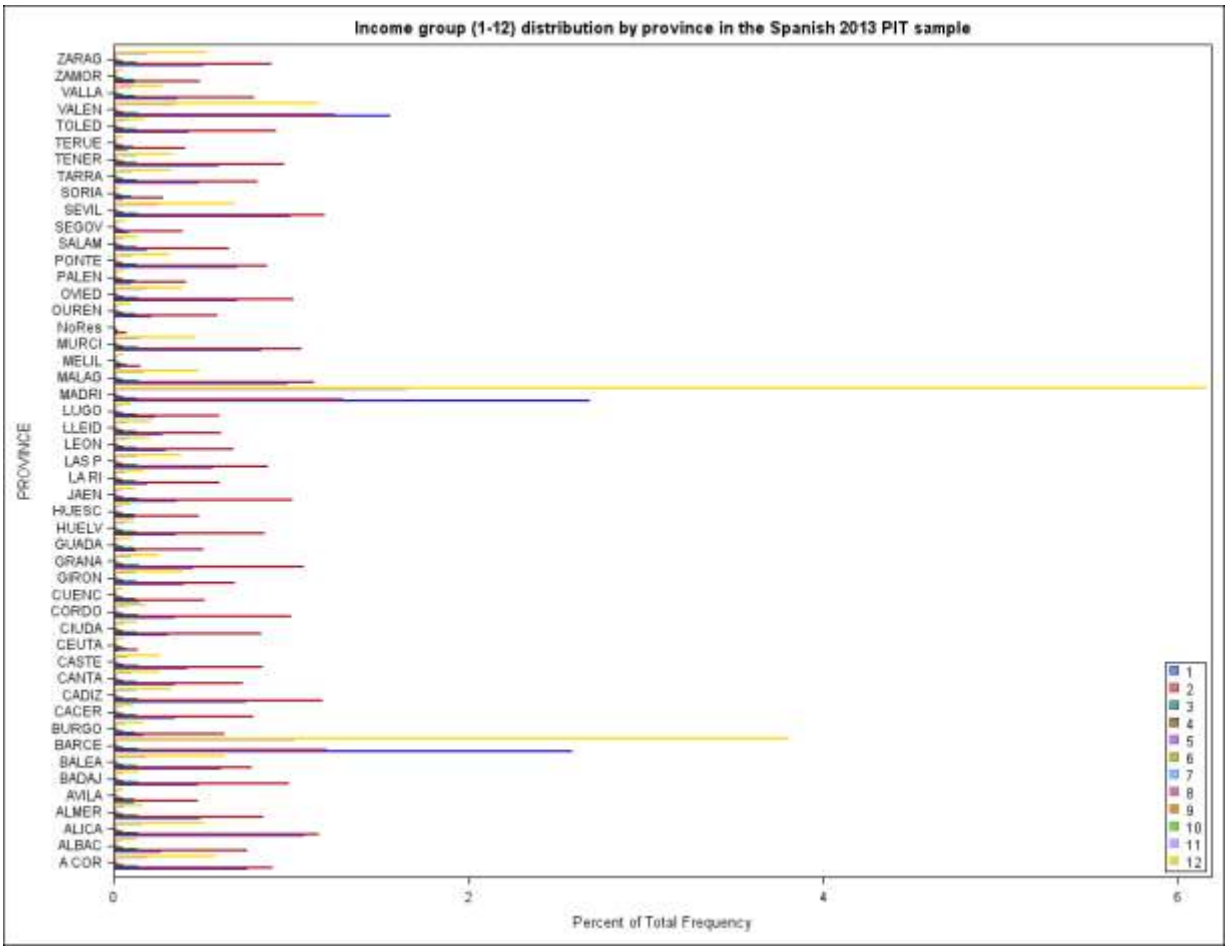
The SAS System

The FREQ Procedure

SEGINCOME	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	517649	23.95	517649	23.95
2	799940	37.01	1317589	60.95
3	125423	5.80	1443012	66.76
4	49336	2.28	1492348	69.04
5	25711	1.19	1518059	70.23
6	15827	0.73	1533886	70.96
7	9993	0.46	1543879	71.42
8	7339	0.34	1551218	71.76
9	5404	0.25	1556622	72.01
10	4171	0.19	1560793	72.20
11	140902	6.52	1701695	78.72
12	459952	21.28	2161647	100.00

Source: own production using data drawn from the Spanish Personal Income Tax 2013 annual sample.

Graph1: Micro-data sample sizes and their distribution by income group (1-12) and province:



As shown in the graph above, provinces with the highest proportion of low income groups (relative to the total number of tax payers in the province) are Valencia, Alicante, Málaga, Jaén, Cuenca and Zamora.

In contrast, provinces with the highest proportion of high income groups (relative to the total number of tax payers in the province) are Madrid, Barcelona, Baleares, Valencia and Zaragoza.

IV- The problem of identifying tax evaders. Scope and limitations of the study.

The main aim of this research is to explore the applicability of data mining techniques for fraud prediction and detection. In this scope, using historical patterns to identify suspicious behavior similar to known fraud patterns would be ideal, but the lack of information about real fraud records makes this task unapproachable.

Instead, we have built an algorithm that searches data for **anomalies that could indicate fraud or error.** In order to generate our dependent variable, the key data for the construction of this algorithm are the "tax deductions", that is, the expenses declared by tax payers that reduce their taxable income. We calculate those tax deductions as the difference between the *gross income tax payable* and the *net income tax payable*.

The algorithm segments the sample population according to the income group criteria (1-12, Table1) and labels as "potential fraud" those taxpayers who declare tax deductions and reductions that are above 95 percentile of their segment. The results of this criterion are described as follows:

Table2: Potential fraud and no-potential fraud sample sizes in Spanish PIT sample.

The SAS System

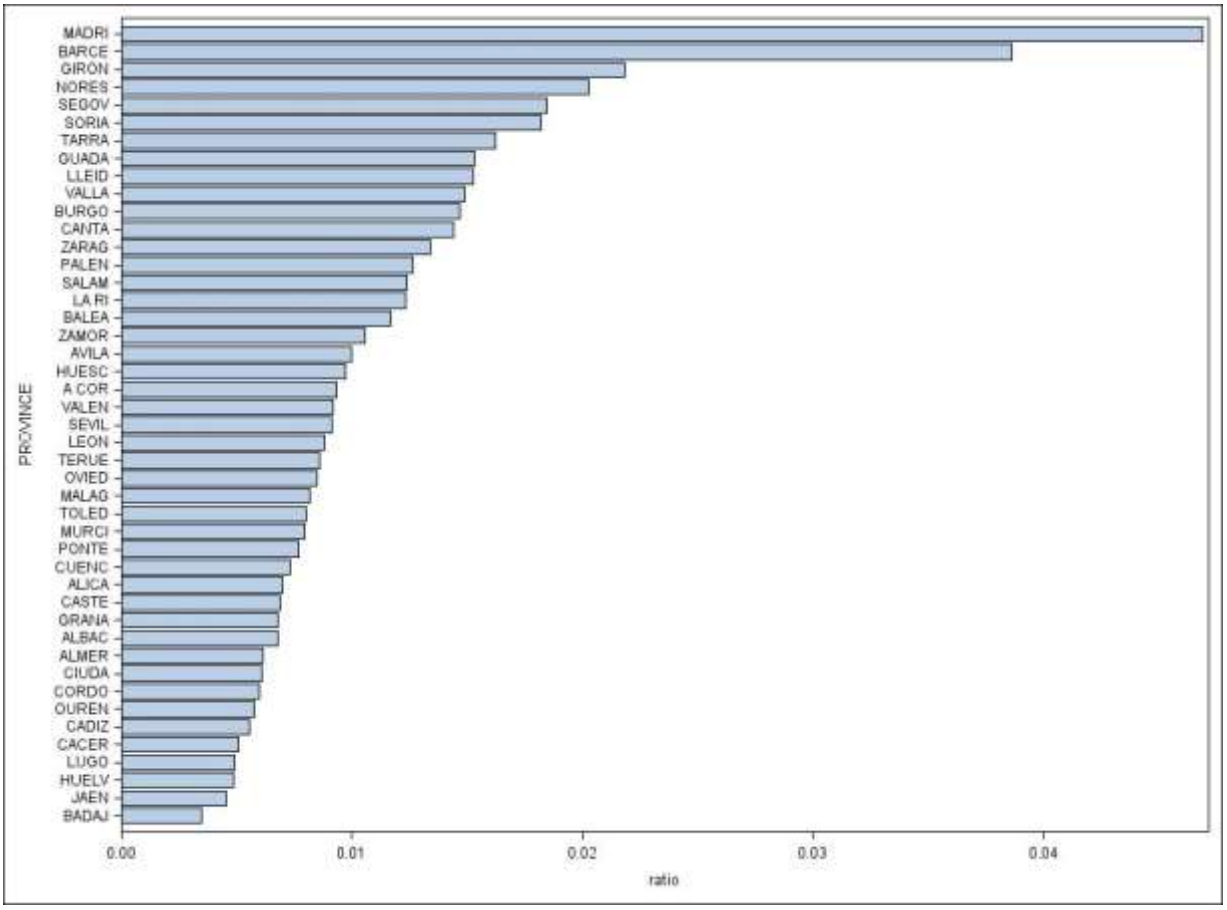
The FREQ Procedure

pf	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	2127061	98.40	2127061	98.40
1	34586	1.60	2161647	100.00

The above information shows that the percentage of potential fraud according to this algorithm is 1.6% (34.586 individuals), so there is a significant imbalance in the proportion of '1' values (potential fraud) in the target variable. This imbalance can lead the model to assign a false '0' (no potential fraud) forecast to **all the taxpayers in the sample** and, consequently, obtain a 98.40% of well classified registers (percentage near to 100% in terms of accuracy) .

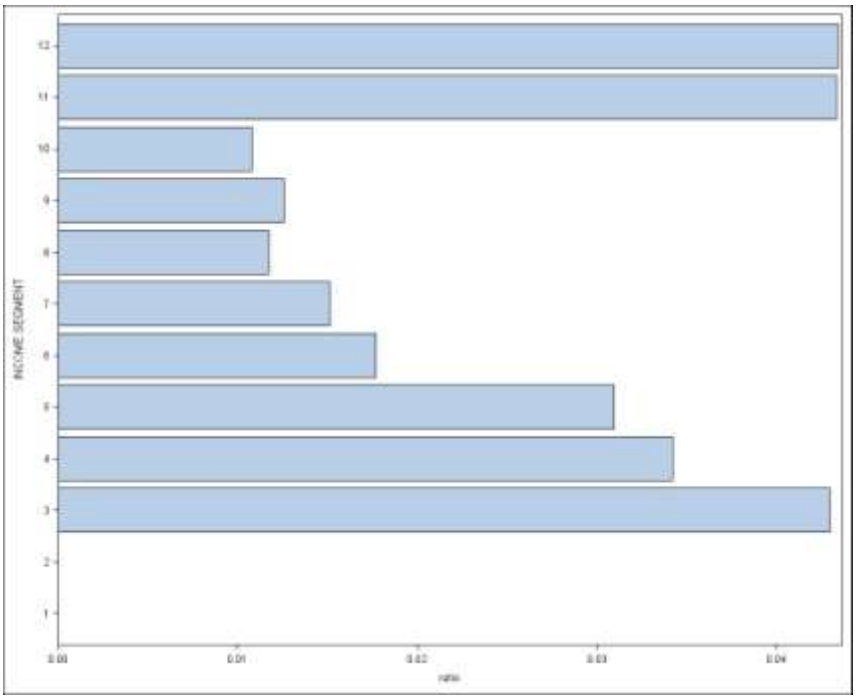
In the next section we propose several approaches to correct the confusion matrix for this imbalance target variable.

Graph2: Potential fraud based on deductions/reductions percentile. Distribution by province (as percentage of province sample size). 2013 Spanish PIT sample



As shown in ***Graph2***, provinces with more proportion of potential fraud relative to their sample size are Madrid, Barcelona and Girona.
Next graph shows the distribution by income segment.

Graph3: Potential fraud based on deductions/reductions percentile. Distribution by income segment (as percentage of income segment sample size). 2013 Spanish PIT sample



From the information provided in ***Graph3***, highest potential fraud is more probable in medium income groups (3 and 4) and upper income groups (11 y 12). Note that income groups 1 and 2 do not include enough taxpayers with deductions for applying a percentile criterion.

V- Estimation Strategy.

We have divided the estimation strategy into two steps. To begin with, we try to find the smallest subset of dimensions that results in an accurate model, that is, the parsimonious solution. For our data set, which contains 469 input variables, it is a necessary task to prevent the model to be over-trained or simply fail to be built, both alternatives possible with large data sets. Despite the fact that we are aware of the great amount of algorithms available for quantifying variable importance, most of these selection processes identify the significance of each variable individually, and skips the opportunity to incorporate the interaction between variables. In most cases, the interaction of two statistically insignificant variables may have a significant effect on the target variable. Even though artificial neural networks and decision trees[1] are still capable of incorporating these effects, the technique that best summarizes the information of all the input variables into ortogonal factors is Principal Components Analysis.

Therefore, we firstly apply a Principal Components analysis as a dimension reduction technique and then we introduce the resulting principal components (that account for just under 80% of the total variance of the initial set of 469 variables) in the input layer of a Multilayer Perceptron. We will now outline the techniques applied.

[1] Building multiple C5 models using different random row samples, which is called the Monte Carlo Method. The variable importance results from each trial are averaged to minimize the variability between single scoring runs.

V. A. Principal Components analysis as a dimension reduction technique.

Given the initial number of variables in the data set in order to effect the characterization and identification of patterns, we first perform a principal component analysis as a dimension reduction technique.

The goal in principal components analysis is to find the minimum number of dimensions that are able to explain the largest variance contained in the initial set of indicators. We intend to simplify the information which gives us the correlation matrix to make it easier to interpret.

Principal component analysis was originated by Pearson (1901) and later developed by Hotelling (1933). The application of principal components is discussed by Rao (1964), Cooley and Lohnes (1971), and Gnanadesikan (1977). Exceptional statistical treatments of principal components are found in Kshirsagar (1972), Morrison (1976), and Mardia, Kent, and Bibby (1979).

Given a data set with p numeric variables, we can compute up to p principal components. Each principal component is a linear combination of the original variables, with coefficients equal to the eigenvectors of the correlation or covariance matrix. The eigenvectors are customarily taken with unit length. The principal components are sorted by descending order of the eigenvalues, which are equal to the variances of the components.

The principal components meet the following properties (Rao 1964; Kshirsagar 1972):

- The eigenvectors are orthogonal, so the principal components represent jointly perpendicular directions through the space of the original variables.

- The principal component scores are jointly uncorrelated. This property ensures the lack of multicollinearity when we use then as input variables in a regression model.

- The first principal component has the largest variance of any unit-length linear combination of the observed variables. The jth principal component has the largest variance of any unit-length linear combination orthogonal to the first j-1 principal components. The last principal component has the smallest variance of any linear combination of the original variables.

- The scores on the first j principal components have the highest possible generalized .Variance of any set of unit-length linear combinations of the original variables.

- The first j principal components provide a least squares solution to the model:

$$Y=XB+E$$

Where:

Y is an nxp matrix of the centered observed variables;

X is the nxj matrix of scores on the first j principal components;
B is the jxp matrix of eigenvectors;
E is an nxp matrix of residuals;

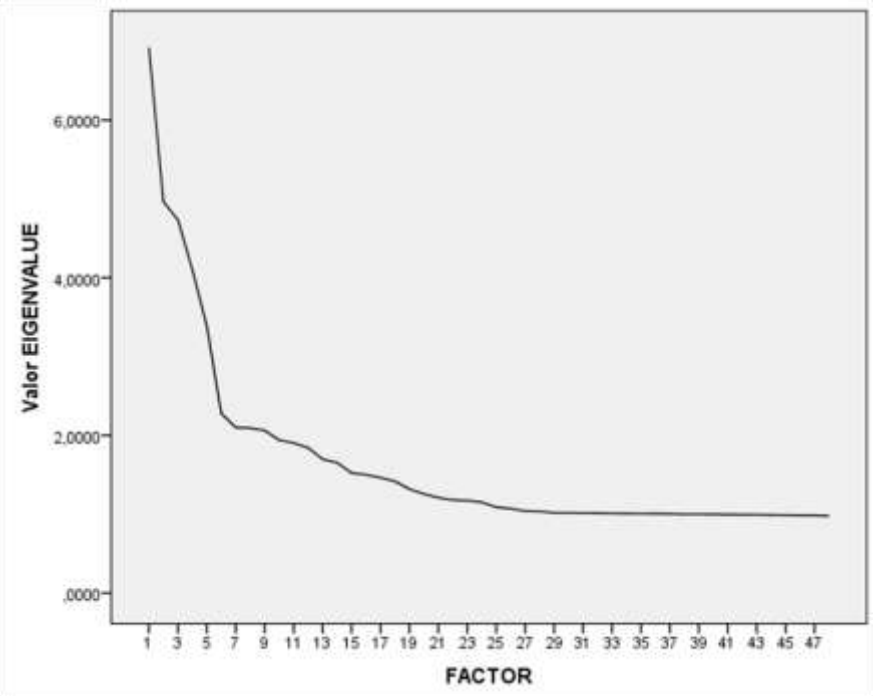
Our goal is to minimize the trace of E'E. That means that the first j principal components are the best linear predictors of the original variables among all possible sets of j variables, although any nonsingular linear transformation of the first j principal components would provide an equally good prediction.

V.B. Principal Components analysis. Experimental results.

Applying PCA to a set of more 469 variables related to the sample, we select the first 48 principal components (Graph5) which are able to explain just under 80% of the total variance of the initial group. We have computed the principal components from the correlation matrix instead of the covariance matrix for two reasons. First, because of the dependency of the covariance matrix on the units of the input variables. Second, the variance differences of the input variables

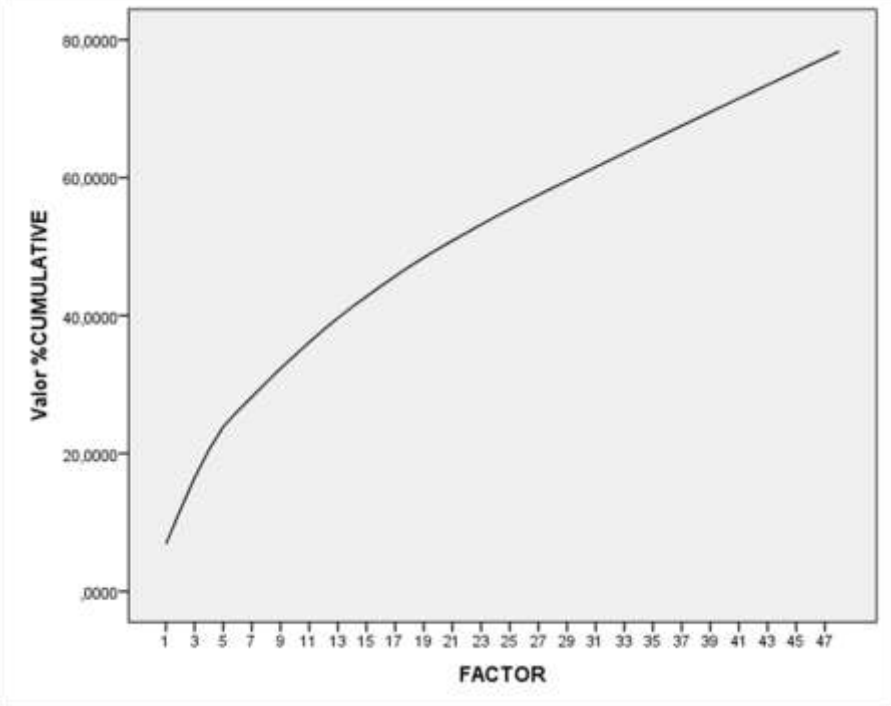
The initial communalities, i.e. the amount of common variance of the variables explained is 1. That is because PCA is based on the assumption that the whole variance can be explained by the factors. In Graph4 and Graph5 we can find the eigenvalue corresponding to each factor and the % cumulative variance explained (scree plot).

Graph4: Eigenvalues table from PCA to a set of 469 variables from 2013 Spanish PIT Sample.



Eigenvectors are determined to transpose the given data. Every eigenvector has an eigenvalue that measures the amount of variance which is in the data in the direction of the eigenvector.

Graph5: Scree Plot. % Cumulative variance explained related to the number of factors.



The Scree Plot (Graph5) visualizes the number of components versus the cumulated percentage of variance explained. The results are based on the unrotated components. As shown in the graph above, with 48 factors we are able to explain just under 80% of the total variance of the group.

V.C. Multilayer Perceptron: some details about the algorithm.

The multilayer perceptron (MLP) is a feed-forward, supervised learning network with up to two hidden layers. It is a function of one or more predictors that minimizes the prediction error of one or more targets. This model uses predictors and targets of both types, continuous and categorical. References to fundamentals of this type of artificial neural network can be found, among others, in Parlos (1994) , Bishop, C. M. (1995),Ripley, B. D. (1996) ,Haykin, S. (1998) andFine, T. L. (1999).

As for the advantages of this network, unlike some other statistical techniques, it does not need to make prior assumptions concerning the data distribution. Another advantage to take into account is that it can model highly non-linear functions and can be trained to accurately generalize when presented with new, unseen data.

We will provide a brief description of the model and the most basic algorithm for training the MLP, known as **backpropagation**. This technique, used for solving the problem of error minimization in MLP, was first stated by D.E, Rumelhart, C. E. Hinton and R. J. Williams[2].

To be more precise, the MLP consists of a system of simple interconnected neurons (or nodes). We represent in graph7 the non-linear mapping between the input vector and the output vector in the 2013 Spanish PIT sample. In our model, there is one hidden layer which consists of seven neurons. The nodes are connected by weights and output signals, which are a function of the sum of the inputs to the node modified by a simple nonlinear transfer function called "activation function". It is the overlapping of many simple nolinear transfer functions that enables the MLP to approximate extremely non-linear functions.

In this particular case, we use IBM Spss Modeler software implementation of MLP which uses the Hyperbolic Tangent function as activation function for the hidden layers:

$$\gamma(x) \quad = \tanh(x) = \frac{e^{x} - e^{-x}}{\mathrm{e}^{\mathrm{x}} + \mathrm{e}^{-\mathrm{x}}}$$

and the following function for the output layer :

$$\gamma(0) \quad = \frac{1}{1 + \mathrm{e}}$$

$$\gamma(1) \quad = \frac{\mathrm{e}}{1 + \mathrm{e}}$$

as in the output layer we only have the binary variable "potential fraud", that only accounts for 0 or 1 values.

[2]Published in Nature 323, 533-536, 9 October 1986.

The output of a node is scaled by the connecting weight and fed forward to be an input to the nodes in the next layer of the network. This implies a direction of information processing and this is the reason why the MLP is known as a feed fordward neural network.

As shown in graph7, the structure of our MLP consists of three layers of neurons: the input layer, the hidden layer (seven neurons) and the output layer (potential fraud variable). The input layer plays no computational role but merely serves to pass the input vector to the network.

MLP are fully connected, with each node connected to every node in the next and previous layer. Hornik et al. (1989) showed that selecting a suitable set of connecting weights and transfer functions, the MLP can approximate any smooth measurable function between the input and the output vectors.

MLP has the ability to learn through training. The objetive of the training phase is to adjust the weights in the network until the desired input-output mapping occurs. Therefore the MLP learns in a supervised manner. During training the output from the MLP for a given input vector may not equal the desired output. An error signal is defined as the diference between the desired and the actual output.We use the Cross-Entropy function to define this error signal for our model.

Training uses the magnitude of this error signal to determine to what extent the weights in the network should be adjusted so that the overall error of the MLP is reduced. The training of aMLP is the procedure by which the values of the individual weights are adjusted so the resulting error is minimized.

The problem of estimating the weights consists in the following parts:

a) Initializing the weights: we take a random sample and apply the algorithm[3] to derive the initial weights.

   The algorithm is used to break out of the local minimum that trainning finds by perturbing the local minimun a certain number of times. If the break-out is successful, the algorithm sets a better initial weight for the next training. The objetive is to find a global minimum by repeating this procedure. As the procedure is expensive for large data sets it is only used on a random sample.

b) Computing the derivative of the error function with respect to the weights . This is solved via the **error propagation algorithm**.

c) Updating the stimated weights. This is solved by the **gradient descent**. The backpropagation training algorithm (Rumerhart et al.,1986) is the most computationally straightforward

[3]The exact procedure is called "**Alternated simulated annealing and training"** and it is available at the document IBM SPSS Modeler 18.0 Algorithms Guide ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/en/ModelerApplications.pdf

algorithm for training the MLP and uses the gradient descent technique to locate the absolute minimun of the error function.

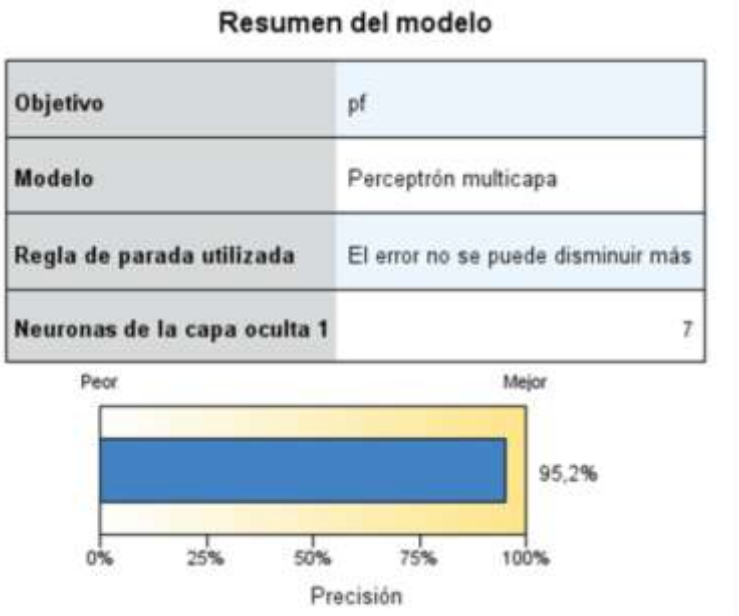
It is important to notice that the error function could contain more than one minimun (local minimum) and it is desirable that the training algorithm does not become trapped in one of these local minimuns. The backpropagation algorithm contains two adjustable parameters,a *learning rate* and a *momentum term*, that can assist the training process in avoiding this problem.

The *learning rate* determines the step size taken during the iterative gradient descent learning process. If the step is too large the network error changes erratically due to large weight changes, with the possibility of jumping over the global minimum. On the contrary, if the learning rate is too small then training will take a long time. The *momentum term* is used to assist the gradient descent process if it becames stuck in a local minimum. It is possible that the weights can escape the local minimum, adding a proportion of the previous weight change to the current weight change.

Training proceeds during at least one complete pass of the data and the search stops according to several stopping rules based on the performance of the MLP.

V.D. Multilayer Perceptron: empirical results.

Graph6: MLP with a hidden layer using the first 48-th principal components as input vector and the potential fraud (pf) as the output vector. Overall performance .



The table above shows that the percentage of accuracy (wrongly calculated) combining PCA and MLP is 95,2% of the sample. As we mentioned in the previous section, this percentage of accuracy is a consequence of the low percentage of records with target values equal to 1 (potential fraud) related to the ‘0’ value (target class imbalance).

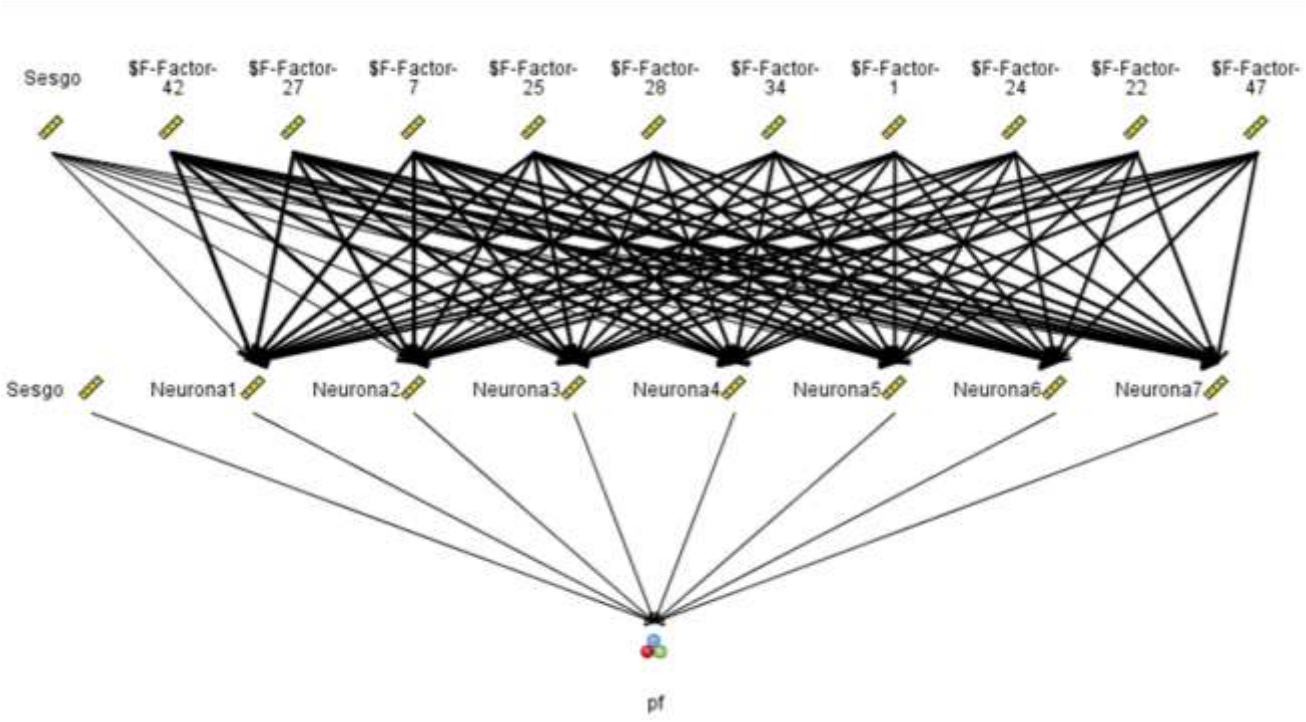
There are several approaches to handle class imbalance: the conventional solution is to resample the data so that the proportions of 1s and 0s are modified. The second one is incorporating priors.

The latest option is based on the assumption that the classification decisions assigned to value '0' is not because the model is not working properly. If we assume no prior probabilities to both values (0 and 1) the threshold to convert the probability to a 0 or 1 label is 0.5. But we can modify this threshold assigning priors to each value and avoiding the model to classify most of the records to the '0' class.

We choose a combination of both techniques to overcome the imbalance target problem. Firstly we under-sample the subset of records assigned to the '0' class (non potential fraud) using the income segment as stratification variable and selecting a random sample of the 20% for each segment[4]. After merging this sample with the subset of records assigned to class '1', the resulting set contains 460000 records with a potential fraud class '1' equal to 7.51% of the total group. In a second step we use priors to correct the initial confusion matrix taking into account the distribution of the class '1' probability previously assigned by MLP (seeMcCormick, K., et al. ,p.255-261, 2013)

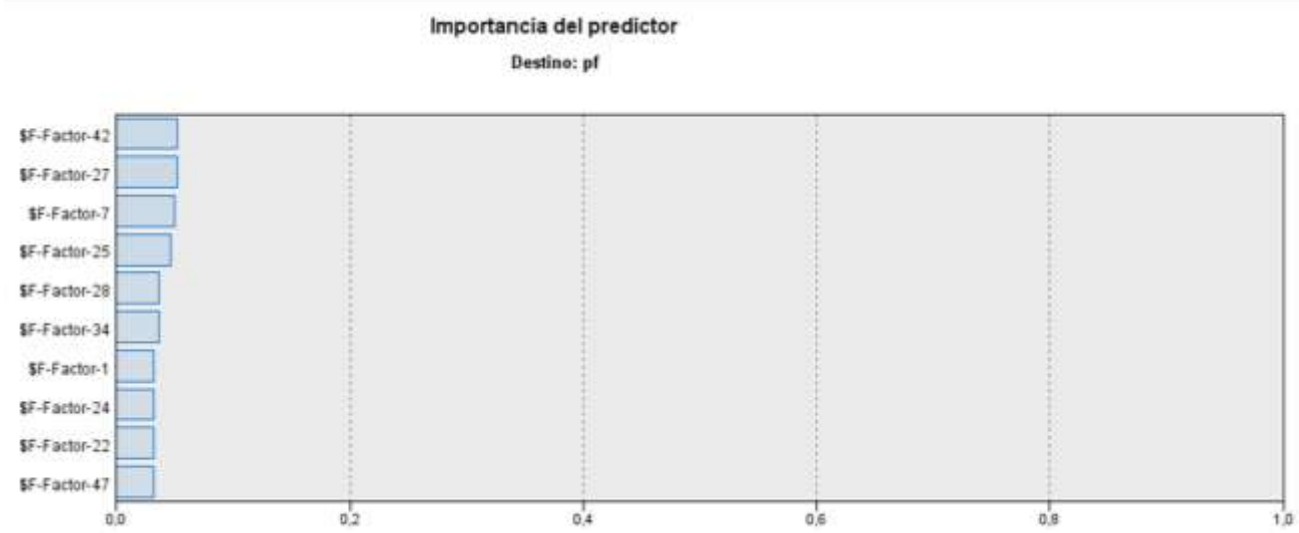
The confusion matrix is usually calculated assigning label '0' to all the records with a probability lower than 0.5 of being in the '1' class. Moving the threshold (equivalent to assigning priors) will help us to provide a real interpretation of the performance of the model.

Graph7: MLP structure graph with a hidden layer using the first 48-th principal components as input vector and the potential fraud (pf) as the output vector. There are 7 neurons in the hidden layer.



[4] Trials with values higher than 20% (30% and 35%) led to very similar results

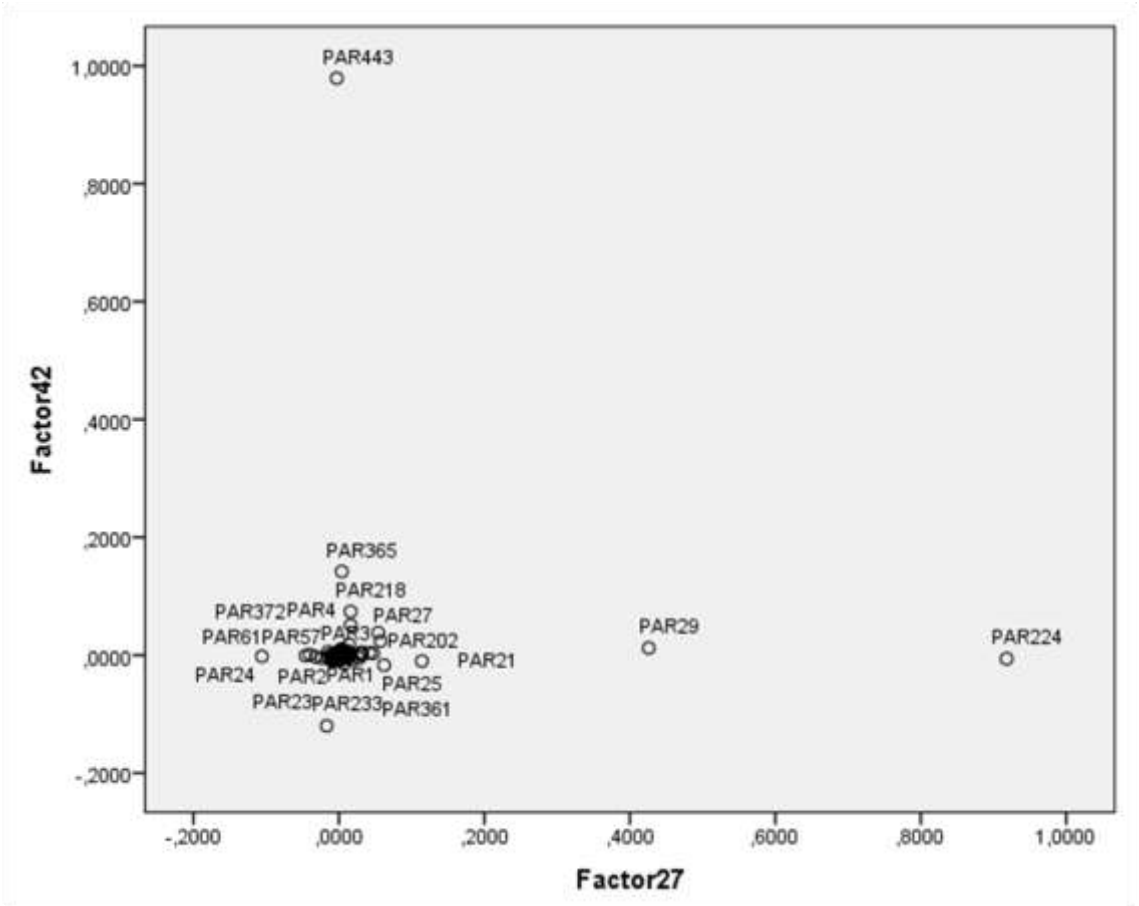
Graph8: Predictors importance in the MLP.



Graph8 shows the predictor importance in the MLP. We see that most important input variables are factors 42, 27, 7, 25, 28, 34, 1, 24, 22 y 47.

The graphical plots of the Varimax-rotated factor loadings show a better association of the inputs variables with the factors. Graphs 9 and 10 illustrate this association with the most important predictors (factors 42, 27, 7 and 25).

Graph9: Rotated Factor Pattern for the most important predictors in the MLP (Varimax rotation). Factor 27 (horizontal axis) and factor 42 (vertical axis).



As we can see, most important predictors (those input variables that scored significantly higher than the others on the axis associated to the factor) related to **factor 42** are **par443** and **par365**.

- The **factor42** is strongly influenced by predictors related to exempted income corresponding to the taxable base on savings.

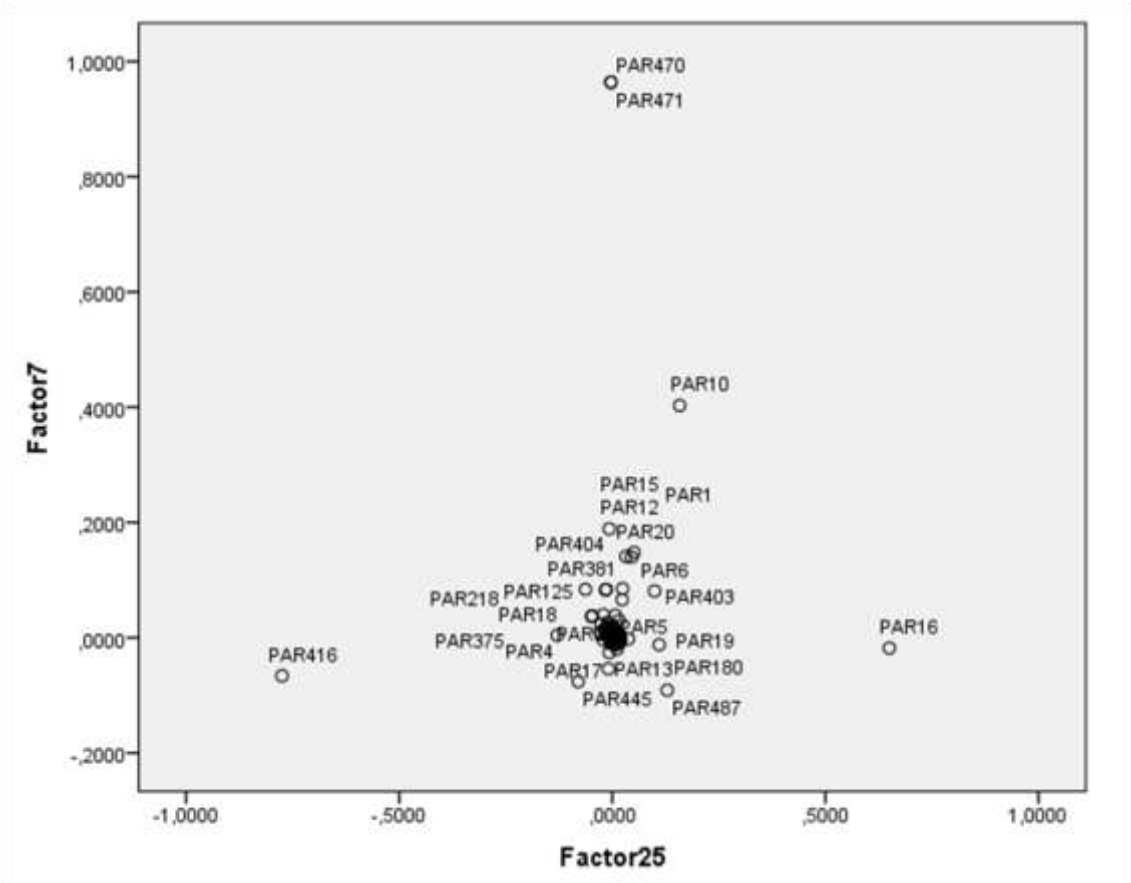
The predictor whose scoring coefficients are bigger on factor 42 is **PAR443** "exempted income corresponding to the taxable base on savings". Although the scoring coefficients of **PAR365** (see Graph9) are not as big as those of the previous variable, **PAR365** does play an important role in the interpretation of this factor, as shown in the graph above. This variable is "*negative net balance resulting from all capital gains and losses corresponding to 2013 treated as taxable income, with the limit of 10% of the net balance of the yields to be included in the general tax base plus income allocations*"

- The **factor 27** is strongly influenced by income from capital gains included in the saving tax base.

As we can observe from **Graph9**, the biggest scoring coefficients in this factor correspond to **PAR224** "*income allocation from entities under the international fiscal transparency system (art.91*

of the Spanish PIT Law)[5] and **PAR29** " *Tax-deductible expenses:deposit and administration of negotiable securities".*

Graph10: Rotated Factor Pattern for most important predictors in the MLP (Varimax rotation). Factor 25 (horizontal axis) and factor 7 (vertical axis) .



- The **factor 7** is strongly associated with **deductions for investments in primary residence and deductions in labor income.**

Most important variables in this factor are *"deductions for investments in primary residence*" **PAR470** (State part) and **PAR471** (Regional part).

- The **factor 25** is strongly associated with reductions in low labor income and with rental deduction.

Most important variables in this factor are **PAR16** *"Reductions for labor income (regulated in art. 20 of the Spanish PIT Law) whose annual net labor income is lower than 14.450 euros"* and **PAR487***" rental deduction : for the rent of the taxpayer's usual residence"*, ***PAR180*** *"net reduced return on agricultural, livestock or forest activities"*,**PAR19** "*additional reductions applied to income from work for active workers who are disabled"*,

The following table summarizes this information and presents the original variables from the Spanish PIT Sample that most influenced the scores obtained by the neural network.

[5] This article of the Spanish PIT Law explains the income allocation from entities not resident in Spain.

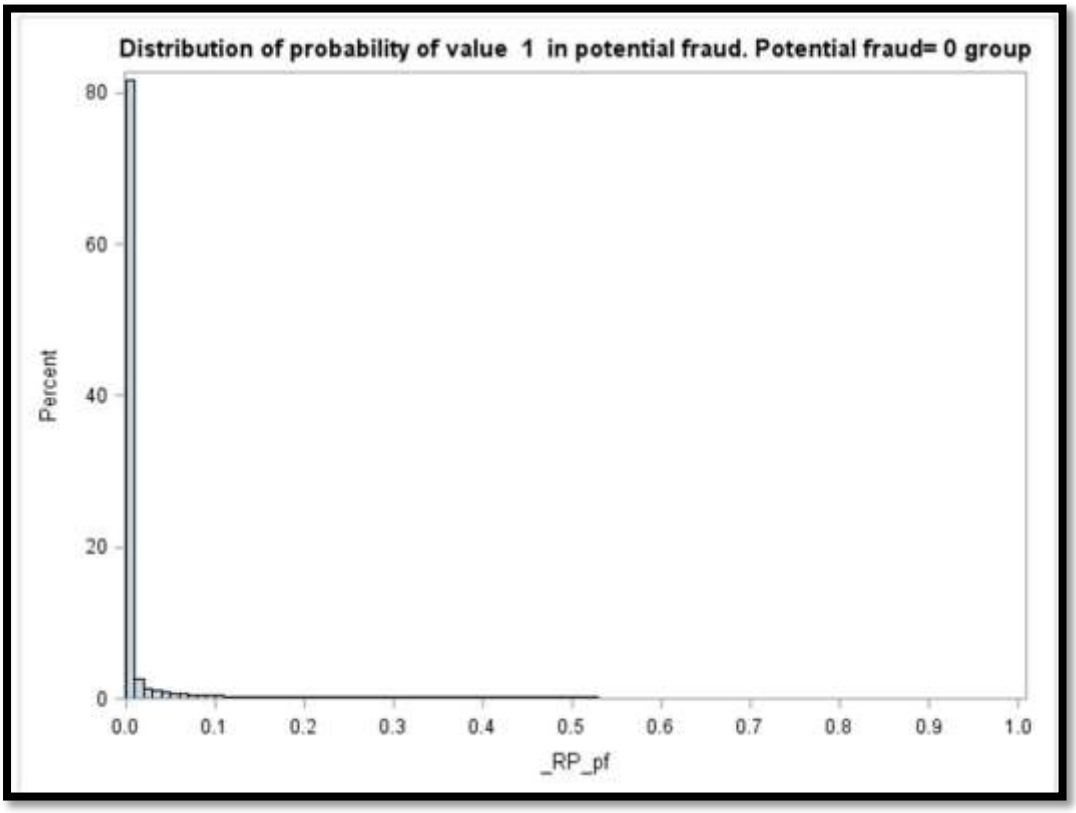
Table3: Most important factors used by the Multilayer Perceptron (in order of importance) and the input variables from the Spanish PIT Sample associated with the factors.

FACTOR 42	Exempted income corresponding to the taxable base on savings.
	Negative net balance resulting from all capital gains and losses corresponding to 2013 treated as taxable income, with the limit of 10% of the net balance of the yields to be included in the general tax base plus income allocations.
FACTOR 27	Income allocation from entities under the international fiscal transparency system (art.91 of the Spanish PIT Law)
	Tax-deductible expenses: deposit and administration of negotiable securities.
FACTOR 7	Deductions for investments in primary residence, State part.
	Deductions for investments in primary residence, Regional part.
FACTOR 25	Reductions for labor income (regulated in art. 20 of the Spanish PIT Law) whose annual net labor income is lower than 14.450 euros.
	Rental deduction : deduction for the rent of the taxpayer's usual residence.
	Net reduced return on agricultural, livestock or forest activities.
	Additional reductions applied to income from work for active workers who are disabled.

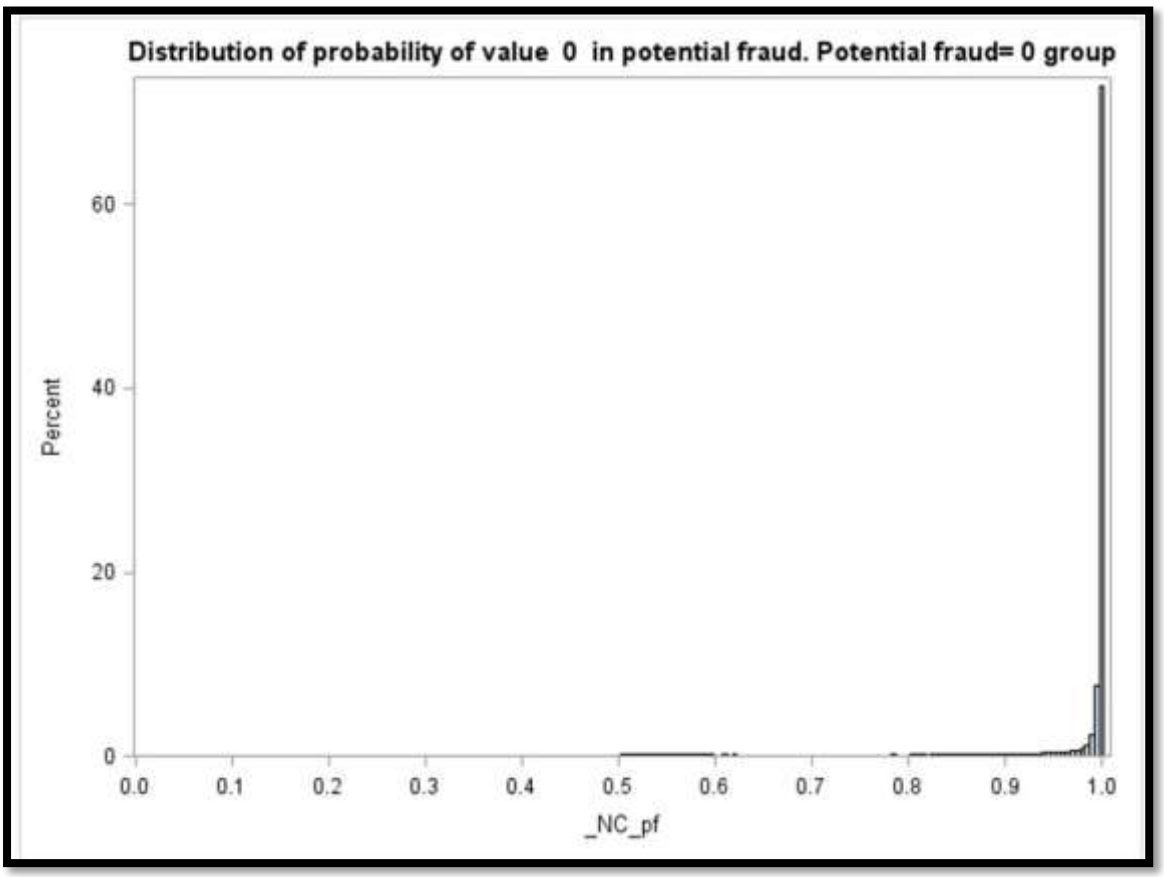
VI-Overall performance of the model. Classification results.

We first analyze the distribution of the scorings assigned to the group of taxpayers with target variable equal to 0 ('non potential fraud' group):

Graph11: Distribution of probability assigned by the model to '1' value for the group with a real '0' value.

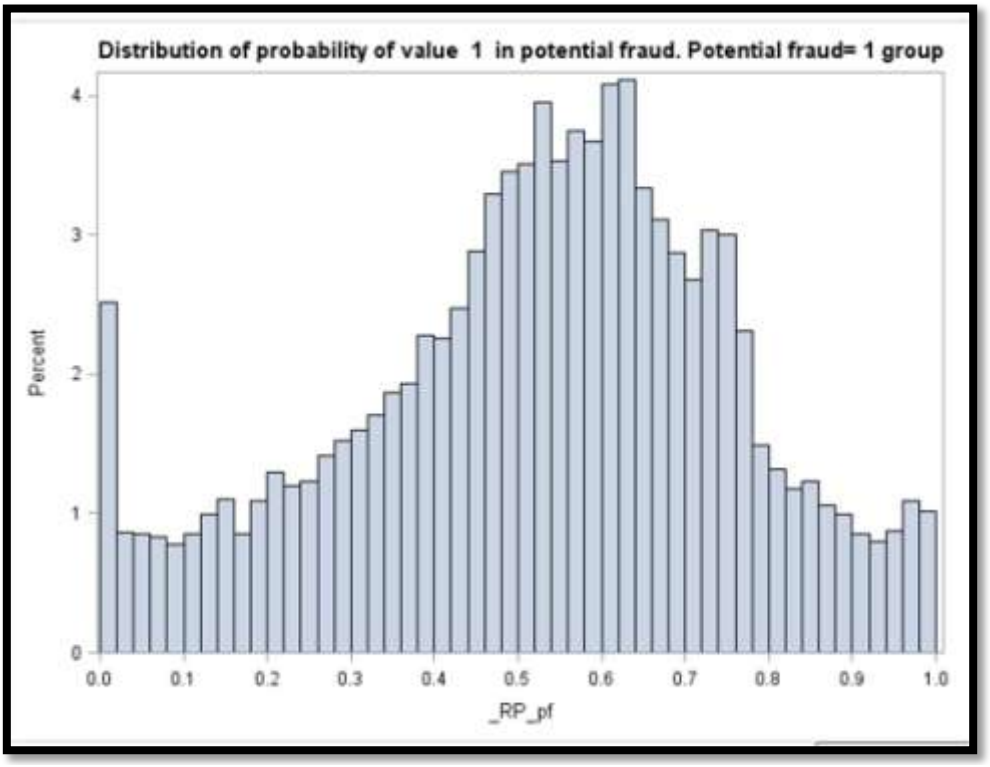


Graph12: Distribution of probability assigned by the model to '0' value for the group with a real '0' value.



From graphs 11 and 12 we can conclude that the performance of the model in the 'non-potential-fraud' group is excelent. From the histrograms above, it is showed that the probability of assigning this group of taxpayers correctly to the "non-potential fraud" group is very high (between 0.8 and 1), and nevertheless the probability of assigning this taxpayers to the "potential fraud" group accumulates in low values (between 0 and 0.2).

Graph13: Distribution of the probability assigned by the model to '1' value of potential fraud for the group with a real '1' .



As we can see, the distribution of the probability of value '1' in the group with potential fraud equal to 1 is less discriminant. As mentioned before, the use of priors consist of assigning a different threshold for labeling the registers to the "potential fraud" or "non-potential fraud" class (for example, we could move the thresold for the probability of potential fraud =1 from 0.5 to 0.08) Those taxpayers with a probability greater than 0.08 of value would be assign to ' potential fraud' group. This could be a good classifier given the distribution of this variable in the other group shown in graph11 and graph12). To better understand how to assign priors, that is, how to choose a threshold for the variables _RP_pf (probability of potential fraud=1) and _NC_pf (probability of potential fraud=0) we follow McCormick, K.,et al (2013) in the following section.

VI.A- Using classifications trees to explore the predictions of a neural network.

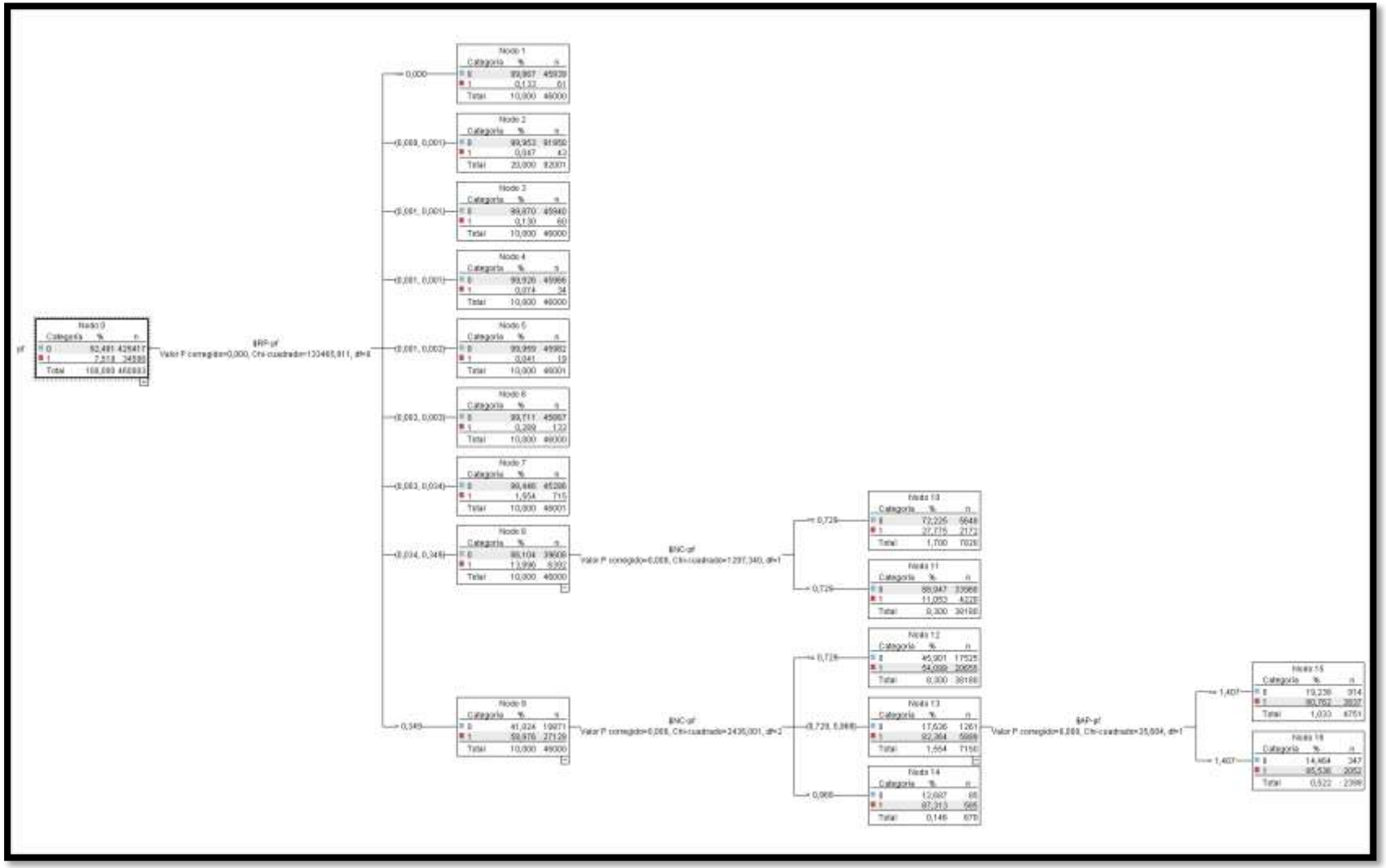
In this section we follow McCormick, K.,et al (2013) using a decision tree algorithm to obtain a population segmentation based on the probability of potential fraud obtained by the MLP. Although MLP are strong performers and we have analyzed the original input variables that have higher weights within the factors that are more important in the MLP, this kind of neural networks do not present an easily-understandable model.

There are various implementations of decision trees we could use to segment the population in terms of the scores produced by MLP. These algorithms mainly differ in the splitting mechanism, that is, the method of finding the optimal partition and the number of new nodes that can be grown from a single node. We use a CHAID algorithm (Chi-square automated interaction detection; (Kass,

1980; Applied Statistics) and (Biggs et al., 1991; Journal of Applied Statistics)) for this task. The CHAID uses the Chi-square independence test to decide on the splitting rule for each node and allows splitting into more than two subgroups. Notice that, the Chi square test is only applicable to categorical data and therefore requires the discretization of all numerical input variables. For each input variable, the classes are merged into a "upper-class", based on their statistical similarity, and manteined if they are statistically dissimilar. These “upper class” variables are then compared to the potential fraud variable for dependency using the Chi-square independence test. The one with the highest significance is then selected as the splitting criteria for the node.

The following graph shows the CHAID structure for our analysis.

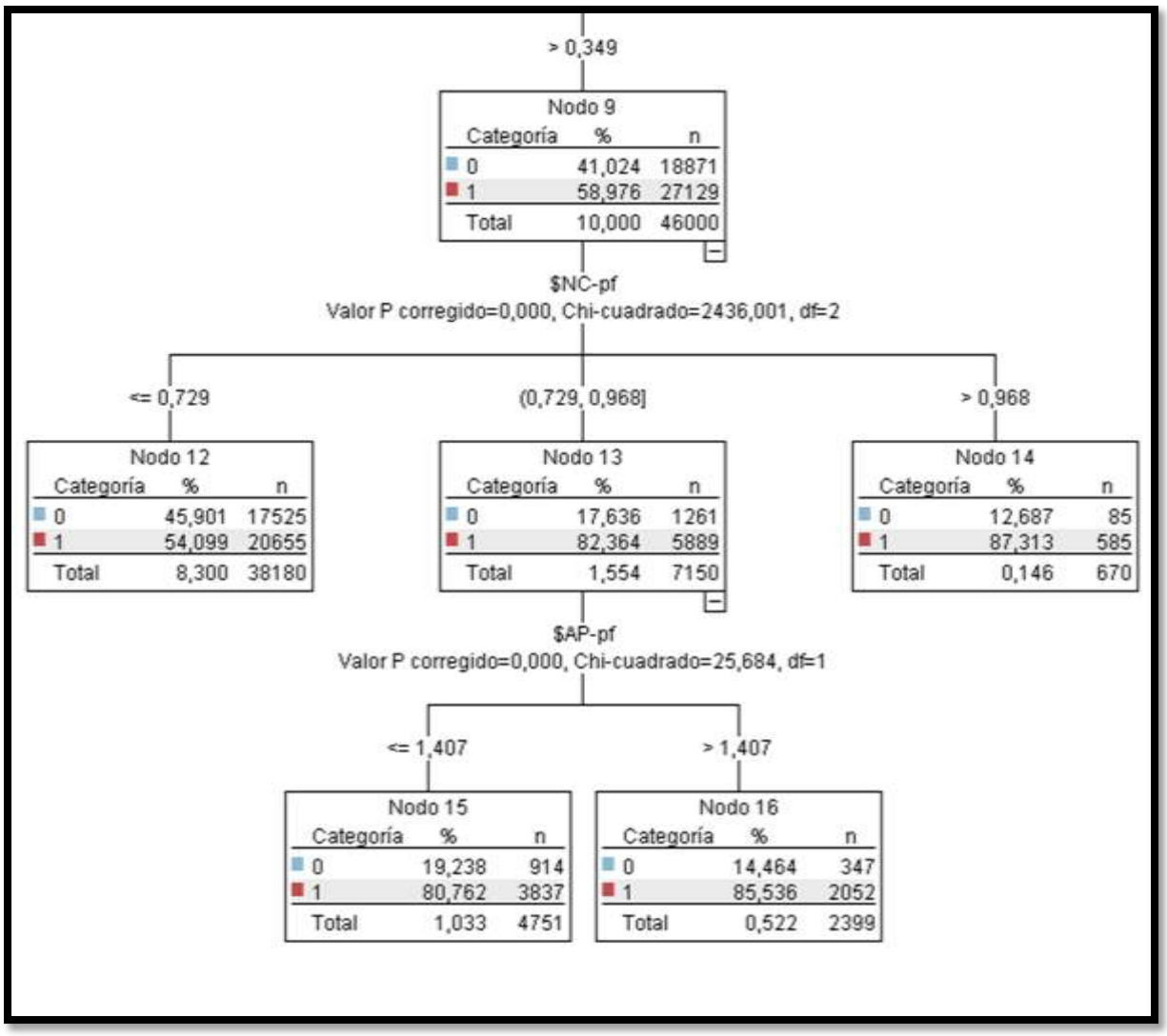
Graph14: Segmentation of the 2013 Spanish PIT Sample according to potential fraud - probabilities assigned by MLP. [6]



[6] NC_PF is the probability of '0' value in potential fraud assigned by MLP.
RP_PF is the probability of '1' value in potential fraud assigned by MLP.

We are especially interested in nodes 14, 15 and 16 whose potential fraud percentages are between 80 and 87%. From now on, the name of the population segments will be that of the nodes of the CHAID tree that generated them.

Graph15: Subpart of CHAID structure :when probability of potential fraud assigned by MLP is greater than the thresold 0.349. Segmentation of the 2013 Spanish PIT Sample according to potential fraud probabilities assigned by MLP. [7]



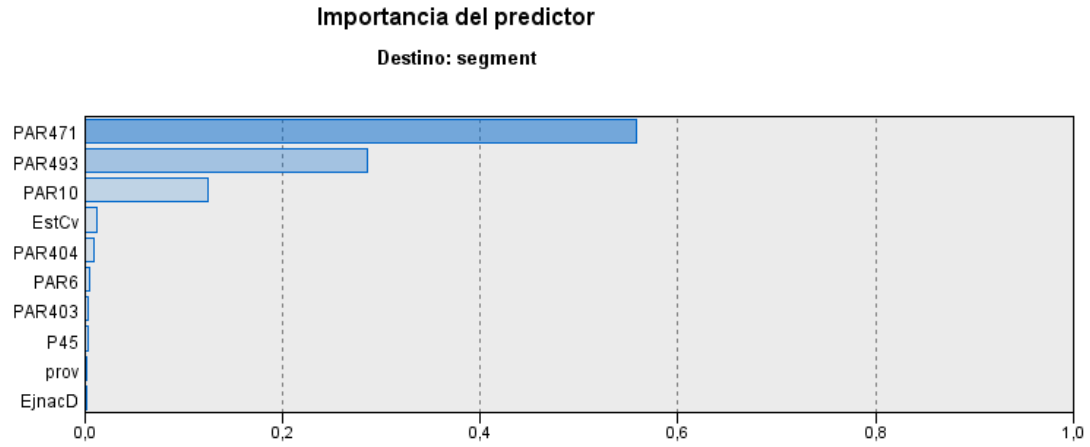
Next subsection explores the characterization of these population segments with high percentage potential fraud records, nodes 14, 15 and 16 (potencial fraud percentage between 80 and 87%) with respect to those which, on the contrary, have a low potential fraud percentage (less than 0.2% : segments 5 and 6).

VI.B- Discriminant input variables that characterize population segments.

The following table summarizes the results explained in this subsection.

[7] NC_PF is the probability of '0' value in potential fraud assigned by MLP.
RP_PF is the probability of '1' value in potential fraud assigned by MLP.

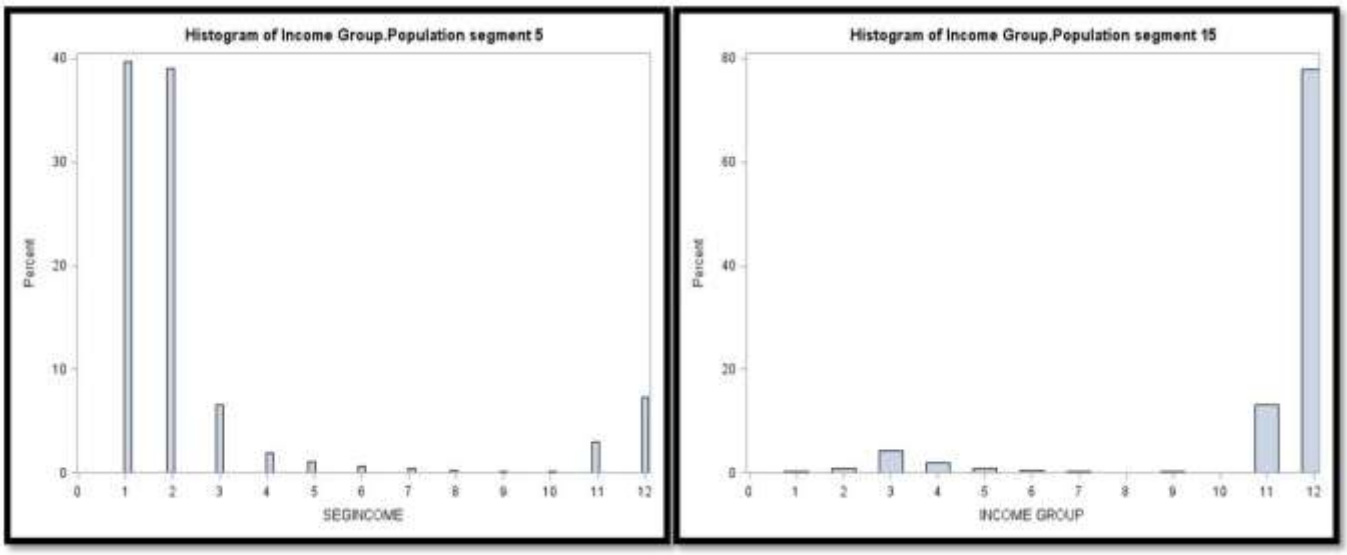
Table4: Most important regressors detected by IBM Spss Modeler using the Chi-square test to discriminate population segments (in order of importance).



We will show the differences in income, province, age, marital status and deductions for investments in main residence, which are the most influential regressors in our model .We will focus on low percentage potential fraud segments (5 and 6) compared to high percentage potential fraud segments (14, 15 and 16).

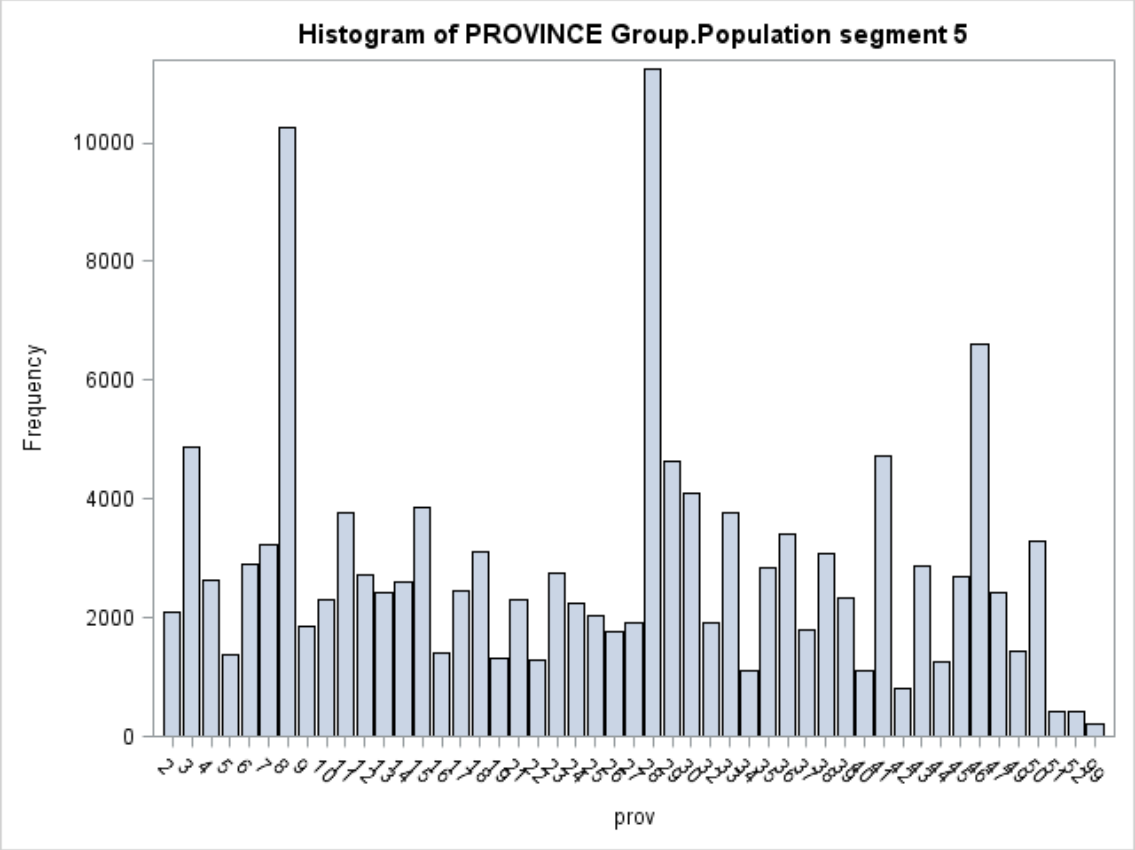
One of most discriminant variables is the **income group.** Segments 5 and 6 with low percentage records of potential fraud (less than 0.2%) show more tax payers in low income groups.

Graph16: Histrogram table of income groups in segment 5 left (99.9% of non potencial fraud) compared to segment 15 right (80% of records with potential fraud).



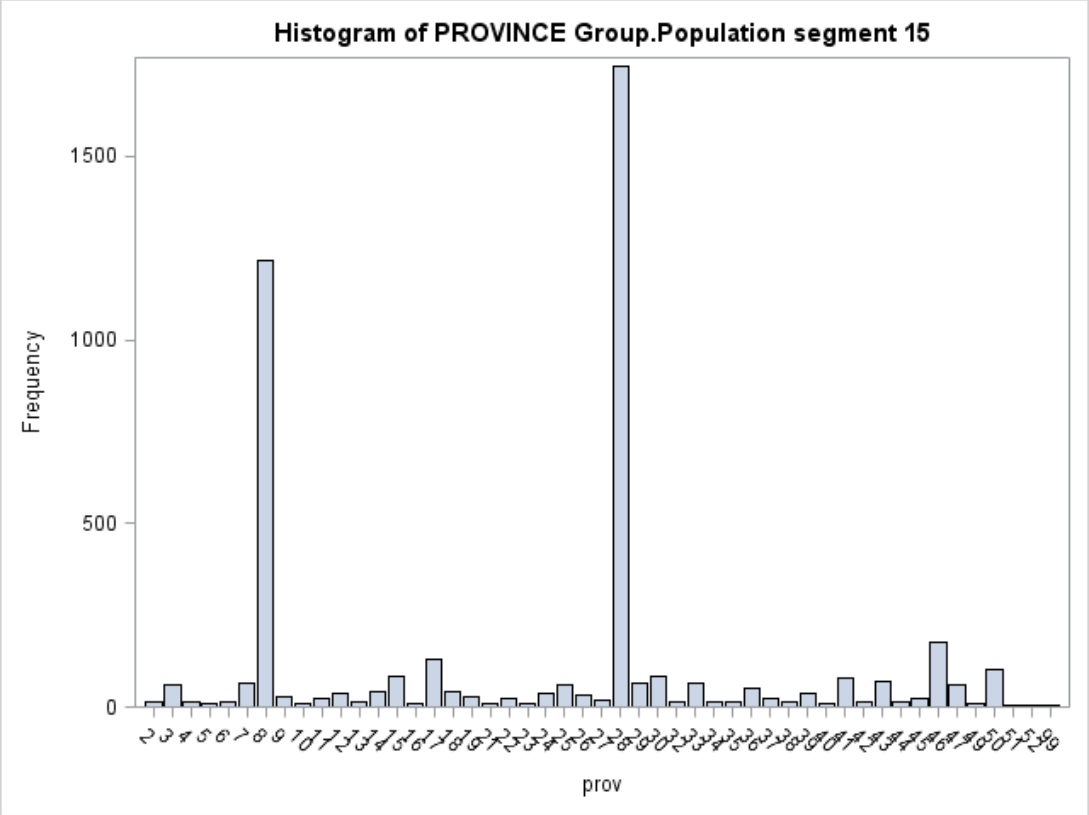
The distribution of the income group variable in this segment 5 is similar to segment 6 and, in general, for the taxpayer segments with low percentage of potential fraud records. The frequencies accumulate in the lower income groups. On the contrary, in population segments with high percentage of potential fraud records (nodes 14,15,16), frequencies accumulate in higher income groups.

Graph17: Bar chart table of province in segment 5 (99.9% of non potencial fraud).



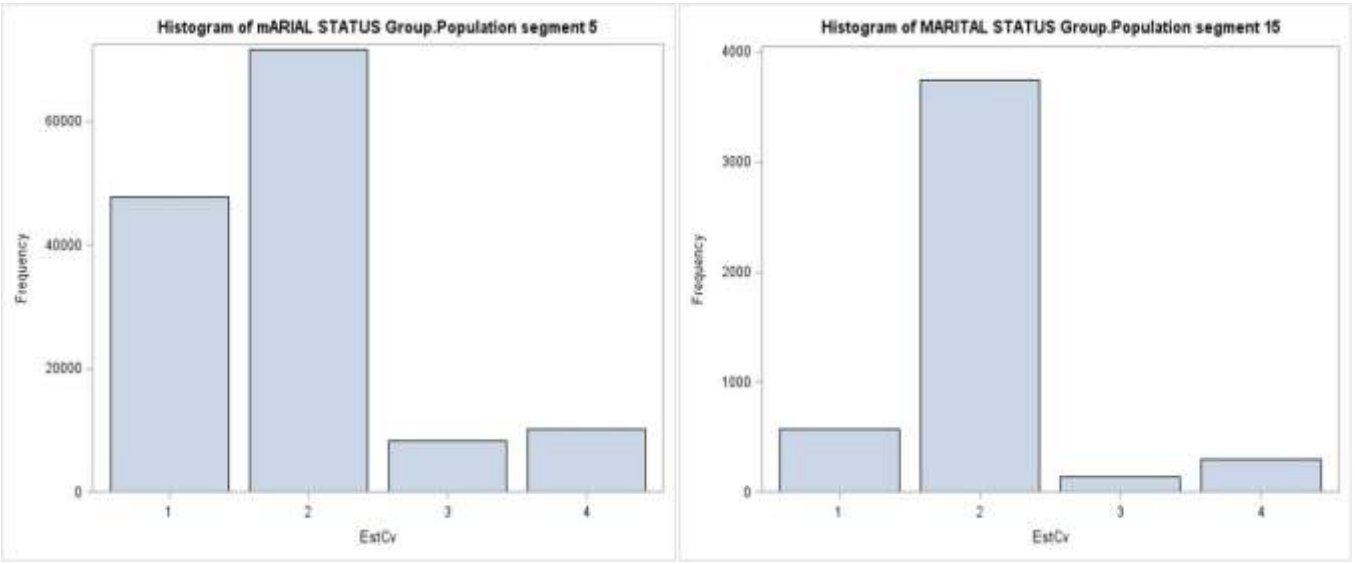
High frequency provinces in this segment of taxpayers with low percentage of potential fraud are Madrid, Barcelona, Valencia , Alicante, Zaragoza and Sevilla. This distribution is similar to the total population of tax payers.

Graph18: Bar chart table of province in segment 15 (80% of potencial fraud taxpayers).



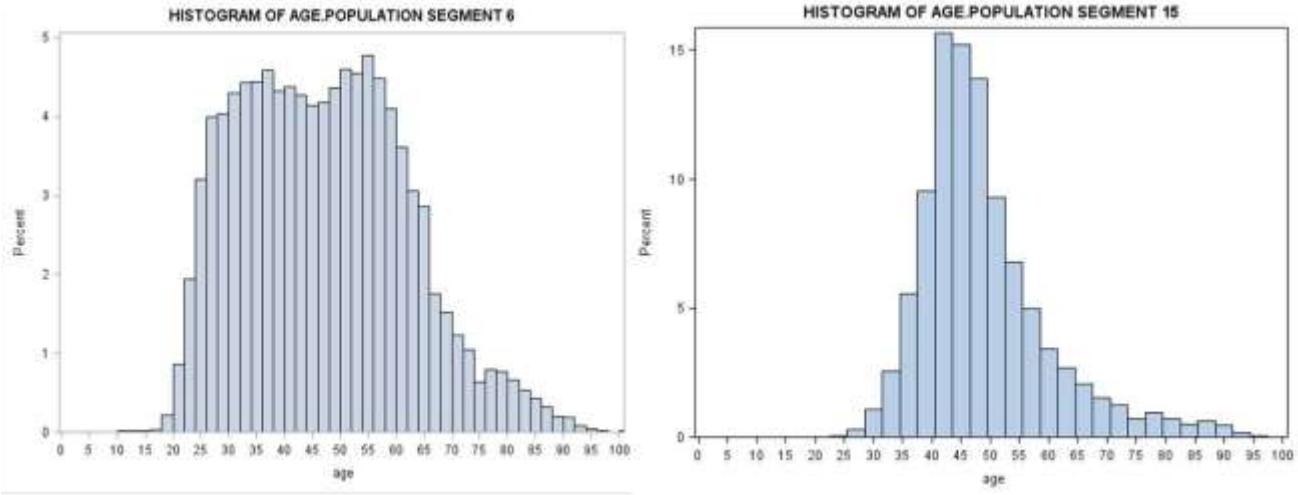
The bins in the histogram corresponding to Madrid (28) and Barcelona(08) accumulate many more frequencies of observations than the previous one (segment 5) The frequency distribution of province variable is more biased about Madrid and Barcelona provinces in this segment (80% of potential fraud).

Graph19: Bar chart table of marital status in segment 5 (left) (99.9% of non potencial fraud) and segmet 15 (right) (80% of potencial fraud taxpayers).



The single category (Marital Status '1') is much less frequent in the high percentage potential fraud segment compared to segments 5 and 6. This distribution is similar to those segments with high percentage potential fraud population

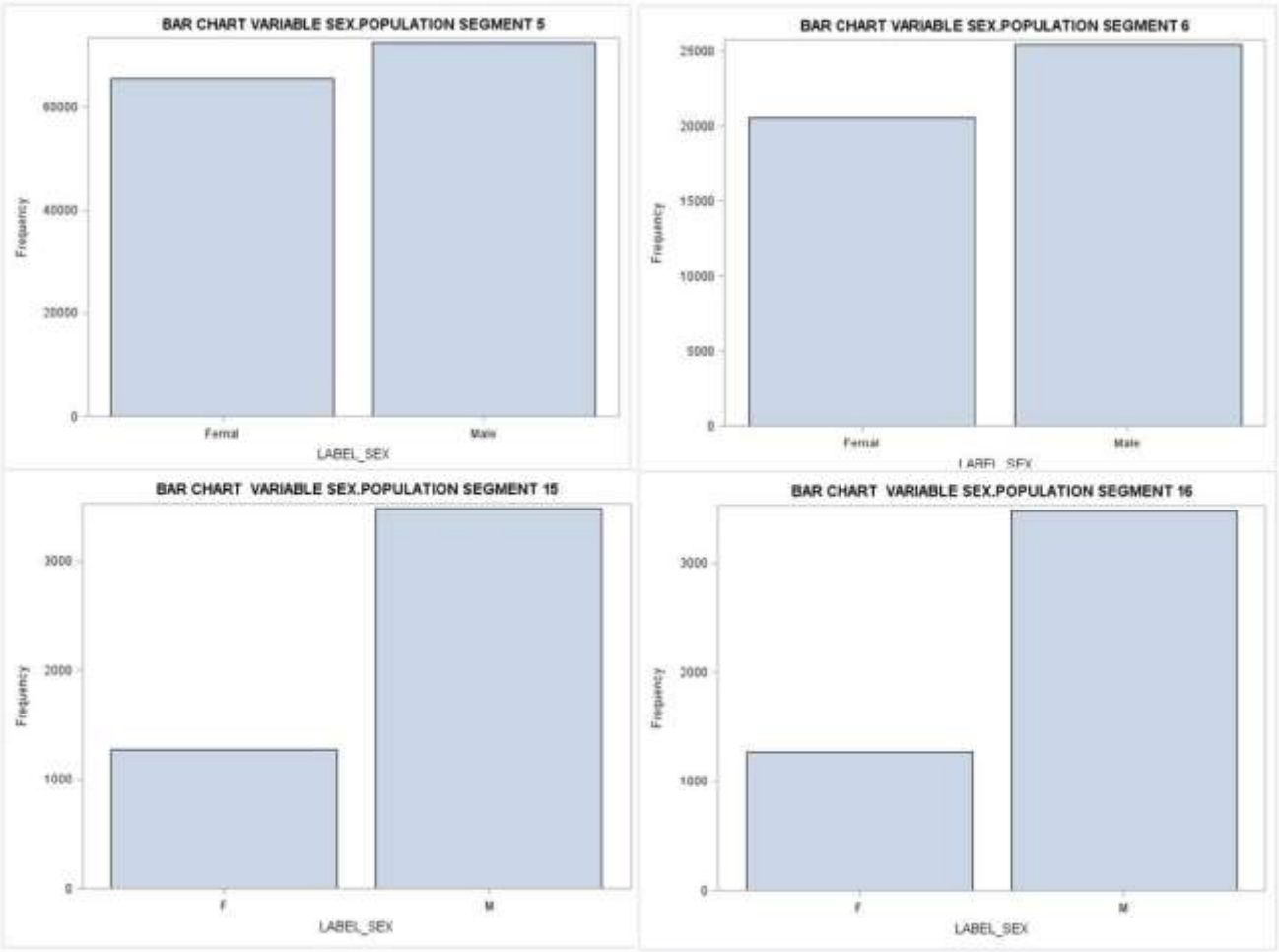
Graph20. Histogram of age in segment 6 (left) (99% of non potencial fraud taxpayers) compared to segment 15 (right) (80% of potencial fraud taxpayers)



The Kurtosis of the distribution in segment 15 [8]leads to focus the attention on population with ages between 40 and 55 years old in 2013.

[8] The shape of the histrogram in segments 14 and 16 are similar, with a high number of records in the 40-55 gap.

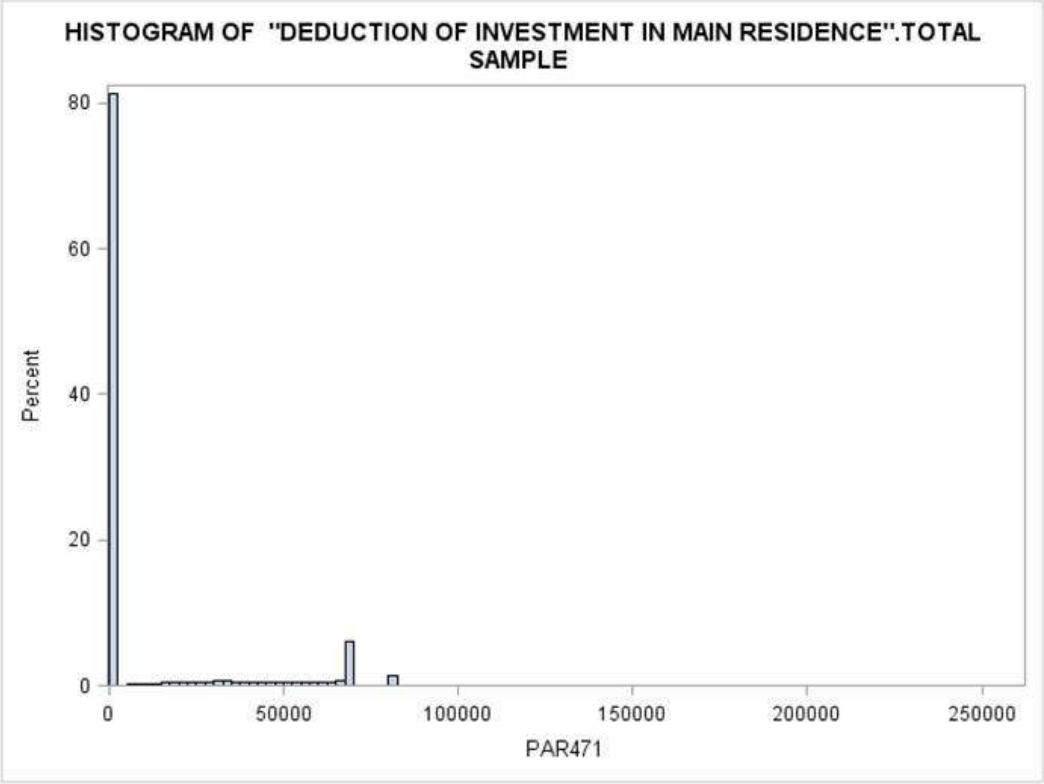
Graph21: Bar chart table of gender. First row: segments 5 and 6 (99.9% and of non potencial fraud). Second row: segments 15 and 16 (80% and 85% of potencial fraud taxpayers)



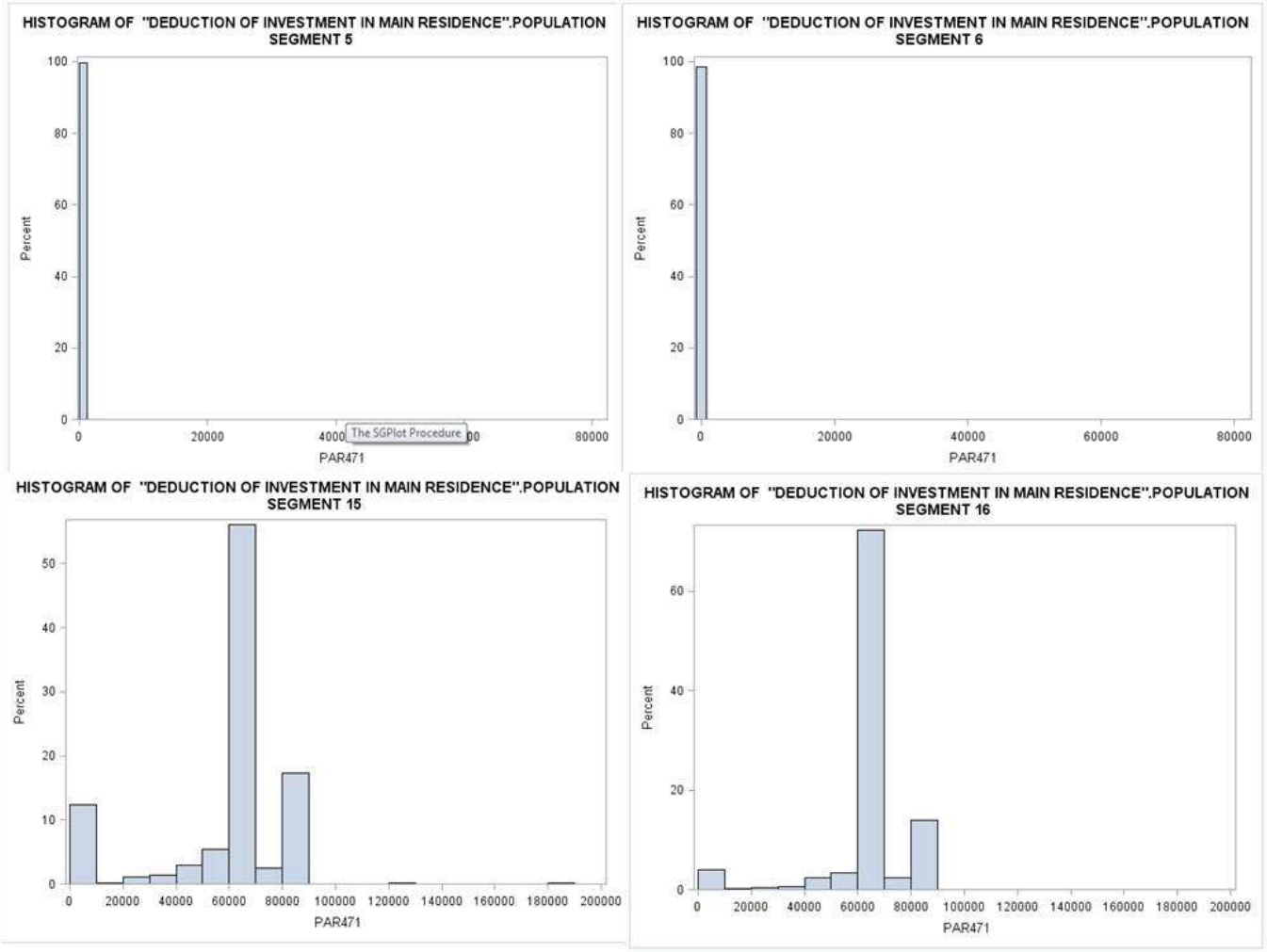
The graph above shows that the distribution of frequencies between the two categories of sex variable is more balanced in segments with low percentage of potential fraud (5 y 6, first row). On the contrary, in segments with high percentage of potential fraud (15, and 16 , second row) the imbalance between men and women is more evident and biased toward men.

The following graph presents the distribution of variable “deductions for investment in main residence” for the total sample. As we can see, this deduction affects only 20% of the total population.

Graph22: Histogram of deductions for investment in main residence. Total population. Figures in cents of euro.



Graph23: Histogram of deductions for investment in main residence. First row: segments 5 and 6 (99.9% and of non potencial fraud). Second row: segments 15 and 16 (80% and 85% of potencial fraud taxpayers). Figures in cents of euro.



Both segments with low percentage of potential fraud (5 and 6, first row) accumulate all the records in the ‘0’ bin. On the contrary, for the high potential fraud segments, more than 70% of the taxpayers accounts for deductions of investment in main residence between 600 and 800 euros in 2013, and aproximately 15% accounts for deductions between 800 and 1000 euros.

Graphs 16-23 reveal that high percentage of potential fraud segments are significantly different and accumulates more records in the categories of male, ages between 40-55, married, Madrid and Barcelona, income greater than 54.000 euros (sum of net incomes, imputed income and capital gains and losses greater than 54.000 euros, ie, income groups 11 and 12) and deductions on investment in main residence between 600 and 1000 euros.

Next subsection formalizes differences between segments using non-parametric test to contrasts population homogeneity in financial variables.

VI.B.1- Testing for differences between segments in financial variables: a non-parametric approach.

In this section we formalize that there are significant differences in the distribution of certain input financial variables depending on the segment of tax-payers. This is equivalent to testing whether two samples (segments of tax-payers) are pulled from the same population. The first point to keep in mind is that most of these financial variables contained in the Spanish PIT sample (and analyzed in previous sections) do not follow a Normal distribution, and therefore, the application of any parametric approach (the T-Student approach for example, where the distribution functions were assumed to be known Normal) for this purpose is open to very serious objections. This is the reason why we are going to use a non-parametric approach.

The Mann–Whitney U test (also called Wilcoxon rank-sum test, or Wilcoxon–Mann–Whitney test) is a nonparametric test of the null hypothesis that it is equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample. As a non-parametric test, it does not require the assumption of Normal distribution and it is nearly as efficient as the T-test on Normal distributions.

Wilcoxon scores are the ranks of the observations, which can be written as:

$$a(R_j) = R_j$$

Where R_j is the rank of observation j, and $a(R_j)$ is the score of observation j.

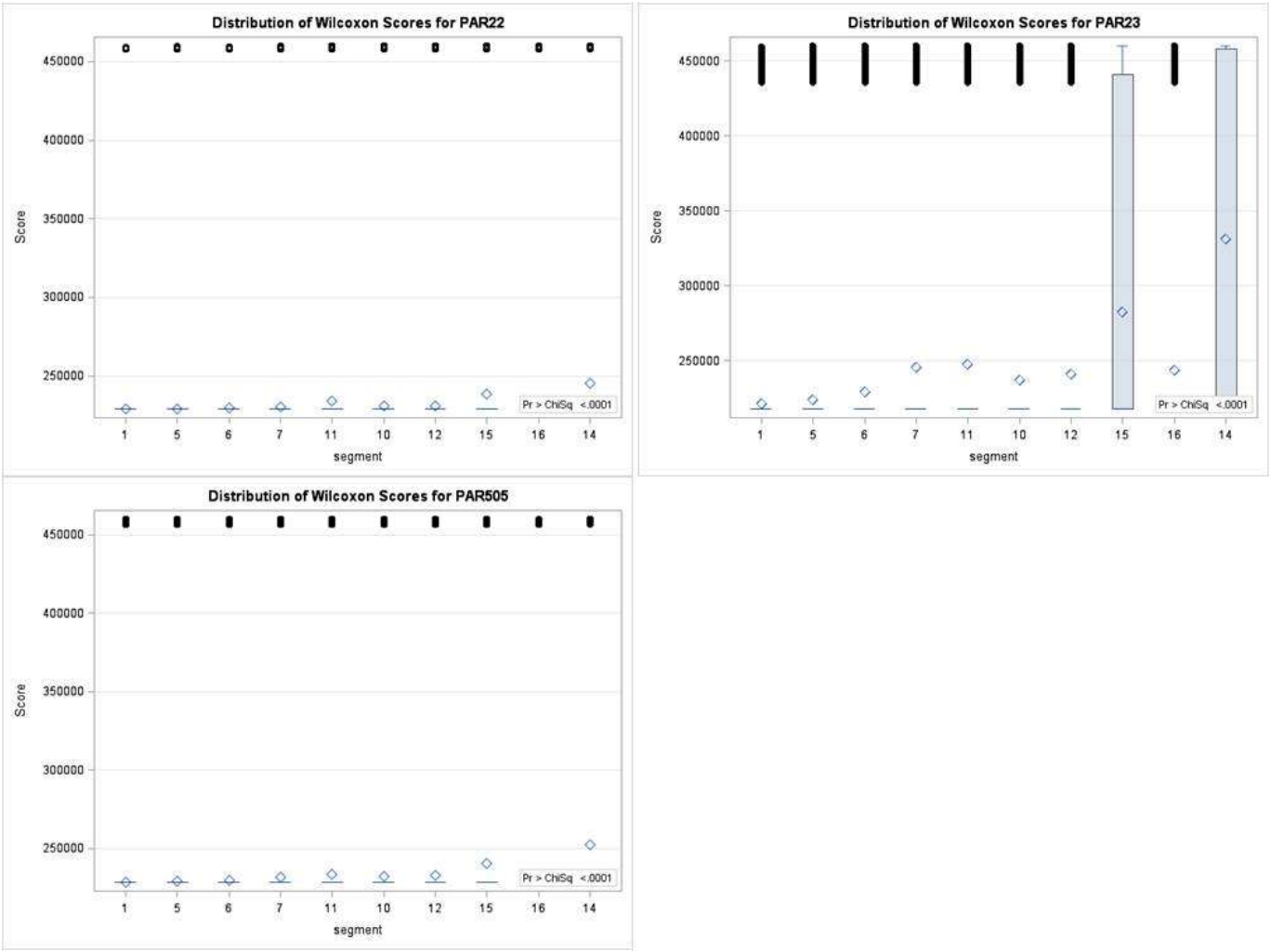
Using Wilcoxon scores in the linear rank statistic for two-sample data produces the rank sum statistic of the Mann-Whitney-Wilcoxon.

The following table contains those financial variables that showed significant differences between segments.

Variable	Label
PAR22	Interest on financial assets entitled to the bonus provided for the 11th Provisional Regulation of the Corporate Income Tax Law
PAR23	Dividends and other income from participation in equity of entities
PAR505	Deduction for double international taxation, due to the income obtained and taxed abroad

Next graph shows the distribution of Wilcoxon Scores for these variables.

Graph24: Distribution of Wilcoxon scores for the following variables: PAR22 "Interest on financial assets entitled to the bonus provided by the 11th Provisional Regulation of the Corporate Income Tax Law" PAR23 "Dividends and other income from participation in equity of entities" and PAR505" "Deduction for double international taxation, due to the income obtained and taxed abroad" .



As seen in the distribution of the scores, segments with high percentage of potential fraud have higher means in these three variables. This result makes us focus attention on those tax-payers with investments abroad who have these investments as sources of income and take advantage of deductions for double international taxation, due to the income obtained and taxed abroad.

VII-Conclusions

Predicting potential fraud is a challenging and daunting task, due to the complexity of the patterns involved and the size of the data sets. In this study, a combined model is proposed to explore the probability of "potential fraud", a binary target variable built from the difference between the *gross income tax payable* and *the net income tax payable* for every income segment.

One of the main problems of this approach is the large number of predictor variables which are multicolinear in nature. In our particular case, we have used 469 variables related to the Spanish Personal Income Tax in 2013. Applying Principal Components Analysis as a dimension reduction technique we have been able to obtain 48 orthogonal regressors that best summarize 80% of the total variance of the group. These principal components are the independent variables that are taken into account for predicting the probability of potential fraud using the Multilayer Perceptron (MLP) model. With the challenge of providing more information related to the initial input variables and their relationship with the MLP scores, we have used the probabilities obtained by the MLP as input variables for a CHAID algorithm, to produce a tax-payers segmentation in terms of potential fraud percentage.

The significant imbalance in the proportions of the target variable (1.6% of records with target value equal to '1' of the total in the sample) has generated problems in the confusion matrix of the MLP model , as the classification model predicted at first wrongly all values equal to '0' . Using a combination of under-sampling methods in the ‘non potential fraud” group and incorporating priors, ie, varying the threshold of 0.5 in the probability that the record belongs to the non- potential fraud class, has help to overcome this problem.

The empirical results showed that most important input variables (deductible expenses and reductions in the tax base) used by the Multilayer Perceptron (in order of importance) to assign potential fraud scores are:

- Exempted income corresponding to the taxable base on savings.
- Negative net balance resulting from all capital gains and losses corresponding to 2013 treated as taxable income, with the limit of 10% of the net balance of the yields to be included in the general tax base plus income allocations.
- Income allocation from entities under the international fiscal transparency system (art.91 of the Spanish PIT Law)
- Tax-deductible expenses: deposit and administration of negotiable securities.
- Deductions for investments in main residence,
- Reductions for labor income (regulated in art. 20 of the Spanish PIT Law) whose annual net labor income is lower than 14.450 euros.
- Rental deduction: deduction for the rent of the taxpayer's usual residence.
- Net reduced return on agricultural, livestock or forest activities.
- Additional reductions applied to income from work for active workers who are disabled.

Characterization of segments reveals that high percentage of potential fraud groups are significantly different and accumulates more tax-payers in the categories of male, ages between 40-55, married, live in Madrid and Barcelona, their income is greater than 54.000 euros (calculated as the sum of net incomes, imputed income and capital gains and losses) and deductions on investment in main residence are between 600 and 1000 euros.

In addition, a non- parametric analysis of financial input variables recommends focusing attention on tax-payers with investments abroad who have these investments as sources of income and take advantage of deductions for double international taxation, due to the income obtained and taxed abroad.

The empirical results reveal that the proposed approach is a promising alternate to potential fraud identification.

REFERENCES

- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, 3rd ed. Oxford: Oxford UniversityPress.
- Bonchi, F., & Giannotti, F. (1999). A Classification-based Methodology for Planning Audit Strategies in Fraud Detection. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, *2*(4), 303-314.
- DANIEL, M. (2013). Application of Data Mining Technology to Support Fraud Protection: The Case of Ethiopian Revenue and Custom Authority.
- Fine, T. L. (1999). *Feedforward Neural Network Methodology*, 3rd ed. New York: Springer-Verlag
- Fox, W. F., Luna, L., & Schaur, G. (2014). Destination taxation and evasion: Evidence from US inter-state commodity flows. Journal of Accounting and Economics, 57(1), 43-57.
- Gardner,M.W. , & Dorling, S.R. (1998). Artificial neural networks (the multilayer perceptron)- a review of applications in the atmospheric sciences. Atmospheric environment, 32 (14) 2627-2636.
- González, P. C., & Velásquez, J. D. (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. Expert Systems with Applications, 40(5), 1427-1436.
- Gupta, M., & Nagadevara, V. (2007). Audit selection strategy for improving tax compliance–Application of data mining techniques. In *Foundations of Risk-Based Audits. Proceedings of the eleventh International Conference on e-Governance, Hyderabad, India, December* (pp. 28-30).
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*, 2nd ed. New York: MacmillanCollege Publishing
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359-366.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychologicalmeasurement*, *20*(1), 141-151.
- McCormick, K., Abbott, D., Brown, M. S., Khabaza, T., & Mutchler, S. R. (2013). *IBM SPSS modelercookbook*. Packt Publishing.
- Micci-Barreca, D., & Ramachandran, S. (2004). Improving tax administration with data mining. White paper. Elite Analytics LLC.
- Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *DecisionSupportSystems*, *50*(3), 559-569.

- Picos Sánchez, F. (2014). Microdatos fiscales en España: caracterización y aplicaciones prácticas.

- Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. Decision Support Systems, 50(2), 491-500.

- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge UniversityPress.

- Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). Parallel distributed processing: Explorations in the microstructures of cognition. Volume 1: Foundations.

- Wu, R. S., Ou, C. S., Lin, H. Y., Chang, S. I., & Yen, D. C. (2012). Using data mining technique to enhance tax evasion detection performance. Expert Systems with Applications, 39(10), 8769-8777.

Detecting Anomalous Pattern-of-Life from Human Trajectory Data

Yazan Qarout and David Lowe

Aston University, Mathematics Department,
Aston Triangle, B4 7ET Birmingham, UK
{qarouty,d.lowe}@aston.ac.uk

Abstract. In smart city data analytics, the observation of human behavioural characteristics from automated sensor analysis can be used to identify sub-clusters and anomalous activity relative to the population's pattern-of-life dynamic. This paper explores this idea using GPS trajectory data obtained from consumer devices. By exploiting both dynamical signal models and probabilistic information derived from non Gaussian departures of the stochastic residual signal, we demonstrate the identification and subclustering of anomalous individuals' behaviours. The methodology uses a dissimilarity framework allowing both identification of anomalies and visual analytics of the high dimensional trajectory dynamics.

Keywords: Smart cities, time series modelling, residual analysis, dissimilarity visualisation.

1 INTRODUCTION

Currently, there are 7.6 billion people in the world and approximately 55% live in urban environments. The integration of analytical technology and smart sensor deployment can provide new avenues to explore and reveal human behavioural characteristics in interacting with the urban landscape in order to improve the life experience of the urban citizen.

Recent literature [1,2] analysing human mobility typically uses video surveillance technology. However, this type of data is typically sparse in its coverage, expensive and intrusive, raising concerns, founded or not, around individual privacy. Other studies [3–5] have used GPS trajectory data. This type of data is informative about peoples movement and is often more cost effective than video surveillance, sampling can be higher, is not geographically anchored, is more diverse, and is without many of the privacy concerns of video or accessing sensitive data in the person's mobile phone.

In this paper we investigate population sub-group characterisation based on simple trajectory data. We propose a dynamic-data model combining explicit time series analysis and error distribution modelling using a dissimilarity framework, following the work of [6]. Using the *Geolife* dataset [7] to test the framework, the basic GPS tracks are converted into sliding window features. Model

space dissimilarity metrics are used to topographically visualise intra-group clustering. The approach is augmented by also exploiting the uncertainty which is contained in the distribution of residuals. Through a combination of kernel density estimation and the Jenssen-Shannon divergence to map into the dissimilarity space we increase the robustness of the visualisation results.

Section 2 discusses some related studies. Section 3 explains the details of the proposed framework. Section 4 summarises the algorithm experiments and its results, concluded by section 5.

2 Related Work

Recent approaches to automated citizen profiling through behavioural studies have focussed on various machine learning techniques to derive classification models and direct time series methods. Classification models rely on accurate labelled data; however, typically most data of this nature is unlabelled.

Jiang et al. [5] investigated the pattern of daily clustering of human activities. Using the Chicago Regional Household Travel Inventory (CRHTI) dataset, which is one of the few data sources with fine labelled information. PCA and k-means clustering were used to visualize the daily activity of subjects and cluster their behaviour to reveal socio-demographic information. The clusters segregated the data into appropriately labelled segments using trajectory information. However, the problem of identifying mobility patterns is often unsupervised and the proposed methodology may only be used with supervised data and so is not generalisable.

Another paper by *Shih et al.* [3] used the Geolife dataset [7] to uncover personal mobility patterns from trajectory data with the aim of identifying pre-condition sufferers of Alzheimer's disease (on the basis that a symptom of Alzheimer's patients is spatio-temporal confusion). The algorithm used training data representing 'normal' behaviour then it was exposed to test data. If the similarity between the movement patterns of the training and test data is lower than a pre-set threshold, the test data was considered anomalous. Although the results showed up to 97% accuracy rate in detecting abnormal mobility patterns, there was no ground truth data, and explicitly, no real information on the Alzheimer state of each subject.

Suzuki et al. [4] used a Hidden Markov Model (HMM) to analyse trajectory data. Using Multi-Dimensional Scaling (MDS), the data was projected onto a lower dimensional space, then, by K-means clustering, human motion pattern clusters were acquired. Anomalies were detected by firstly computing the likelihood of the trajectory to the HMM. If the maximum likelihood value is smaller than a predefined threshold, the trajectory is considered anomalous. This method was tested on trajectories extracted from video imagery and has not been demonstrated successfully on GPS data. This was also the case with *Fu et al.* [8] and *Piciarelli et al.* [9] where they used two level hierarchical clustering and Support Vector Machine (SVM) clustering respectively in order to detect anomalies in trajectories from video feeds.

3 Methodology

Our alternative approach combines dynamical and stochastic models of time series unified by a dissimilarity framework. The main stages are (i) Feature extraction, (ii) Time series modelling, (iii) Residual Modelling, (iv) Dissimilarity representation, (v) Visualisation and finally (vi) Anomaly investigation. We consider each of these in turn.

3.1 Feature Extraction

GPS data often contains three main variables of interest: longitude, latitude, and time. These three variables are used to trace a person's movements. A time series of these variables will be called a trajectory. The first step is to calculate the *trajectory descriptors*: change of distance Δd, change in speed Δs, and change in rate of turn Δr. Assuming the trajectory sequence $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, where $\mathbf{x}_t \in \mathbb{R}^3 = (time, longitude, latitude)$, the data is transformed to the sequence $Y = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$ where $\mathbf{y}_t \in \mathbb{R}^4 = (time, \Delta d, \Delta s, \Delta r)$.

Using the Haversine formula [10], the distance between two trajectory points can be calculated. Assume T points with longitude λ_t and latitude τ_t, the distance travelled between two points is

$$h(\frac{\Delta d_t}{r}) = h\left(\Delta\lambda\right) + cos(\lambda_{t-1}).cos(\lambda_t).h\left(\Delta\tau\right), \tag{1}$$

where $h(\theta) = sin^2\left(\frac{\theta}{2}\right)$ and $r = 6371km$ is the radius of the Earth assumed to be constant on all locations on its surface. The change in speed Δs is then calculated by $\frac{\Delta d_t}{\Delta time_t}$. The remaining feature, change in rate of turn Δr, describes the angle at which the persons trajectory turns in comparison with the previous trajectory point:

$$\Delta r_t = arctan\left(\frac{\lambda_t - \lambda_{t-1}}{\tau_t - \tau_{t-1}}\right). \tag{2}$$

GPS data is usually non-uniformly sampled and either over or under sampled making them difficult to analyse by conventional modelling. In order to tackle this problem, we convert to an equivalent set of uniformly sampled time series by passing the data through a sliding window with a fixed time length in order to generate a uniformly sampled window sequence. The window passes through sequence Y, storing a subset of the series within it. This will form a sequence of windows $\mathbf{W} = (W_1, W_2, \ldots, W_M)$, where $W_m = (\mathbf{y}_{t_m}, \ldots, \mathbf{y}_{t_m+L_m-1})$ with cardinality $|W_m| = L_m$ and t_m is the index for the first recorded element in sequence W_m.

The data within each window is then reduced to a 12-dimensional vector as follows: In each window W_m in the window sequence $\mathbf{W}$, the mean μ, variance σ, minimum min, and maximum max of each of the trajectory descriptors Δd, Δs, and Δr are evaluated and chosen to represent that window, replacing the original data.

$$g(W_m) = g(\mathbf{y}_{t_m}, \ldots, \mathbf{y}_{t_m+L_m-1}), \tag{3}$$

$$g(W_m) = g\left(\begin{bmatrix}\Delta d_{t_m}\\ \Delta s_{t_m}\\ \Delta r_{t_m}\end{bmatrix}, \ldots, \begin{bmatrix}\Delta d_{t_m+L_m-1}\\ \Delta s_{t_m+L_m-1}\\ \Delta r_{t_m+L_m-1}\end{bmatrix}\right), \tag{4}$$

where $g(\cdot) = mean, var, min, max(\cdot)$. By applying this function to all W_m, this results in the feature sequence $F = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_M)$ where the feature vector $\mathbf{f}_m = \{\mathbf{i}_m\}$ and $\mathbf{i} = \{\mu_{\Delta d}, \sigma_{\Delta d}, \min_{\Delta d}, \max_{\Delta d}, \mu_{\Delta s}, \sigma_{\Delta s}, \min_{\Delta s}, \max_{\Delta s}, \mu_{\Delta r}, \sigma_{\Delta r}, \min_{\Delta r}, \max_{\Delta r}\}$. These feature sequences may then be modelled as described next. Note, for this paper we will be assuming the three primary descriptors are uncorrelated for simplicity.

3.2 Time-Series Modelling

We assume an underlying and unknown generative process drives each individual in the population. Therefore we seek a time series model which explains the trajectory features. Since we have no prior knowledge on which models are relevant, we use a simple set of AR processes (a different one for each trajectory), and accept that the uncertainty derived from our model ignorance will still be retained in the distribution of the residuals. We deal with the residuals later.

We assume an autoregressive process of a linear time series model. Assuming the time series data sequence $\mathbf{x} = (x_1, x_2, ..., x_T)$ has a mean of zero, an autoregressive model of order p, $(AR(p))$, will predict the value of x_t such that

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + z_t, \tag{5}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)$ are constant weights, and z_t is noise at time t. The order p is determined using the partial autocorrelation function.

The AR coefficients $\boldsymbol{\phi}$ are estimated as:

$$\begin{bmatrix}x_{p+1}\\ x_{p+2}\\ \vdots\\ x_T\end{bmatrix} = \begin{bmatrix}x_p & \ldots & x_1\\ x_{p+1} & \ldots & x_2\\ \vdots & \vdots & \vdots\\ x_{T-1} & \ldots & x_{T-p}\end{bmatrix} \cdot \begin{bmatrix}\phi_1\\ \phi_2\\ \vdots\\ \phi_p\end{bmatrix}, \tag{6}$$

$$\mathbf{x} = X_{lag} \cdot \boldsymbol{\phi}, \tag{7}$$

$$\boldsymbol{\phi} = X_{lag}^{+} \cdot \mathbf{x}, \tag{8}$$

where X_{lag}^{+} is the pseudo-inverse of X_{lag}.

We repeat this process for each feature sequence of each trajectory separately. Thus in principle each trajectory and each person has a different time series model describing their local dynamics. The feature sequence F is converted into a set of AR coefficients $\Phi = \{\boldsymbol{\phi}_i\}$, $\boldsymbol{\phi}_i = (\phi_1, \ldots, \phi_{p_i})$ and p_i is the AR model order of the i^{th} feature in the feature sequence.

3.3 The Time Series Dissimilarity And Cepstral Analysis

Comparing characteristic differences between individuals' behaviours has been reduced to comparing the distance between models in model space, where each model can have different complexity. Therefore, we convert the time domain models into a set of cepstral coefficients c which are estimated for each model. Cepstral coefficients quickly decay to zero for a *stationary* linear model, therefore, only the first few coefficients are necessary to describe the model. Note, we are assuming our observed trajectories are technically stationary. If not the method needs to be modified to approximate the nearest stationary trajectory. We select a fixed number N of cepstral coefficients $\mathbf{c} = (c_1, \ldots, c_N)$ for each model. The cepstral coefficient distance in equation 11 is used to create a Model Proximity matrix **MP** between all trajectories.

Cepstral analysis is commonly used in speech processing. However, studies such as those by *Martin* [11] and *Kalpakis et al.* [12] have shown the benefit of using cepstral coefficients as a dissimilarity measuring tool for time series data. Given a time series x_t, the cepstral domain can be calculated as

$$c(t) = \mathcal{F}^{-1}(log|\mathcal{F}(x_t)|), \tag{9}$$

where $\mathcal{F}(.)$ denotes the Fourier transform and $\mathcal{F}^{-1}(.)$ is the inverse Fourier transform. As demonstrated by *Kalpakis et al.* [12], the cepstral coefficients $c_1, \ldots, c_N$ for an $AR(p)$ model can be derived from the AR coefficients $\phi_1, \ldots, \phi_p$ by

$$c_n = \begin{cases} -\phi_1 & n = 1, \\ -\phi_n - \sum_{m=1}^{n-1} (\frac{n-m}{n}) \phi_m c_{n-m} & 1 < n \leq p, \\ -\sum_{m=1}^{p} (\frac{n-m}{n}) \phi_m c_{n-m} & p < n. \end{cases} \tag{10}$$

Martin [11] proposed the following metric to calculate the distance d between time series models k and l with cepstral coefficients $c_n^{(k)}$ and $c_n^{(l)}$ respectively.

$$d(\mathbf{x}^{(k)}, \mathbf{x}^{(l)}) = \left[\sum_{n=1}^{N} a_n |c_n^{(k)} - c_n^{(l)}|^2 \right]^{\frac{1}{2}}, \tag{11}$$

where a_n is a pre-specified weight. *Martin* [11] used $a_n = n$. Different problems may require different definitions for the weight to enhance separation.

One of the main advantages of using this metric is based on the fact that cepstral coefficients decay rapidly to zero. This means that only the first few coefficients are needed to represent the time series; thus, making it an effective technique at distinguishing time series of different lengths. *Kalpakis et al.* [12] also argue that cepstral analysis has high discriminatory power, and is invariant under basic transformations such as amplitude translations/scaling and time-shifting.

The result of this process are the model space dissimilarity matrices $\{\mathbf{MP_i}\}$ for each feature $\mathbf{i}$. They may then be consolidated into one matrix $\mathbf{M} = \sum_{\mathbf{i}} \mathbf{MP_i}$, where $\mathbf{M}_{k,l} = [d(k,l)]$, summarising the relative separation between the models of all the trajectories.

3.4 Residual Analysis

In addition to model uncertainty, with real data there will also be observational error and intrinsic noise. All such uncertainties are reflected in the distribution of the residuals: the difference between the model predictions and the observed data [6]. Therefore, to complete our modelling framework we also need to exploit the uncertainty in the noise distribution.

Let $\hat{x}_t = E[x_t]$ be the expected value of observation x_t, then the residual may be simply calculated by

$$z_t = x_t - \hat{x}_t. \tag{12}$$

The probability density function may then be estimated for each residual corresponding to models of different time series as

$$\hat{p}(y) = \frac{1}{Th} \sum_{t=1}^{T} K\left(\frac{y - z_t}{h}\right), \tag{13}$$

where $K(\cdot)$ is the Kernel function, y is the point where the density is estimated at, and h is the bandwidth that acts as the smoothing factor for the distribution.

In this paper we choose to evaluate a separate one-dimensinoal pdf for each feature for each person. To compare dissimilarities between pdfs we use the (symmetrised) Jensen-Shannon divergence

$$JSD(\mathbf{p}\|\mathbf{q}) = \frac{1}{2}\left(D_{KL}(\mathbf{p}\|\mathbf{m}) + D_{KL}(\mathbf{q}\|\mathbf{m})\right), \tag{14}$$

where $\mathbf{m} = \frac{1}{2}(\mathbf{p}+\mathbf{q})$, and $D_{KL}(\mathbf{p}\|\mathbf{q})$ is the Kullback Leibler divergence between distributions:

$$D_{KL}(\mathbf{p}\|\mathbf{q}) = \sum_i p(i) \log \frac{p(i)}{q(i)}. \tag{15}$$

Thus, for each feature $\mathbf{i}$, a comparative Residual Proximity matrix $\mathbf{RP_i}$ is evaluated across all individuals' trajectories using the Jensen-Shannon divergence as a measure of dissimilarity. To consolidate this information, the aggregated sum over all features is taken, ie. $\mathbf{R} = \sum_i \mathbf{RP_i}$.

This finally provides the combined dissimilarity matrix:

$$\mathbf{D} = \alpha\mathbf{M} + (1-\alpha)\mathbf{R} \tag{16}$$

where $\alpha \in [0, 1]$ controls the convex combination of the dynamical and stochastic components.

3.5 Visualisation

To visualise the high dimensional information in $\mathbf{MP_i}$ and $\mathbf{D}$, we use Principal Component Analysis (PCA) as a benchmark method and compare with t-distributed Stochastic Neighbour Embedding (tSNE) [13]. Therefore any anomalous trajectories (being anomalous relative to other trajectories) ought to be visualised as a distinct pattern in the visualisation space. Locally Linear Embedding (LLE) [14] and Isomaps [15] are also used with $\mathbf{D}$ in an attempt to preserve the topographic information contained within the matrix.

4 Results

The methodology was validated using the Geolife dataset [7]. This dataset contains 17,621 GPS trajectories recorded from 182 people over a period of 5 years.

The algorithm was tested on 464 trajectories $\{X_1, \ldots, X_{464}\}$ that were recorded from the movements of 9 different people. The calculated trajectory descriptors $\{Y_1, \ldots, Y_{464}\}$ were passed through a sliding window with a length of 3 minutes and an overlap duration of 1 minute. After extracting the feature sets $\{F_1, \ldots, F_{464}\}$, they were approximated using AR models and the cepstral coefficients were calculated. Equation 11 was then used to construct the 12 proximity matrices $\{\mathbf{MP_i}\}$. The metric was tested using weights $a_n = \frac{1}{n},\ a_n = 1,\ a_n = n$ and $a_n = n^2$. It was concluded that $a_n = n^2$ forms the most observable and distinct clusters, therefore $a_n = n^2$ was selected for the distance metric. Figures 1 and 2 depict the respective visualisations for each proximity matrix $\mathbf{MP_i}$ using tSNE and PCA.

In order to understand the clusters, the trajectories were plotted on geolocated maps. For example, the middle cluster in the tSNE visualisation of the 'Speed Maximum' plot contains the trajectories in figure 3. The journeys have a similar outline and mostly contain movements that are on foot. This movement pattern was absent from other clusters in the tSNE visualisation. Moreover, it is interesting to note that trajectories 313.17 and 1254.17, which belong to person 163 and person 153 respectively, are almost identical although they belong to two different people. Person 153 and person 163 may have been together when on this journey, thus explaining the high similarity in the trajectory outline. The algorithm was able to successfully place these journeys in overlapping proximities. Another cluster within the same plot shows a different pattern. The 'C' shaped cluster on the left mostly consists of trajectories similar to those in figure 4. The cluster contains trips on highways and main roads. Vehicles driving on open roads tend to have different speed behaviours to journeys on foot. Therefore, the algorithm correctly segregated the open road trajectories from those in figure 3. As seen in figure 5, the method is also able to differentiate between vehicular trips of different natures. Examining the topmost cluster in the 'Speed Variance' plot shows trajectories that take place near the centres of main cities such as Shanghai and Beijing. Being two of the most congested cities in the world, it is reasonable that these trajectories can be segregated from others using a speed-dependent feature.

Figure 2 shows the PCA plots for each of the features. When compared to the remaining data, outliers in the plots can be depictions of anomalous trips. For example, figure 6 shows some of the outliers observed in the 'Speed Variance' PCA plot. Although they may seem similar to the journeys in figure 3, these trajectories have subtle differences causing them to be labelled as anomalous. Trajectories start with small movements that appear to be walking/cycling, however, the final points in the journey show significant alterations in speed. This may have been caused by transition into means of faster transportation, or due to errors in the data recording. The alterations in speed also occurs for a short distance on the map. This is interesting since using a fast form of transport to

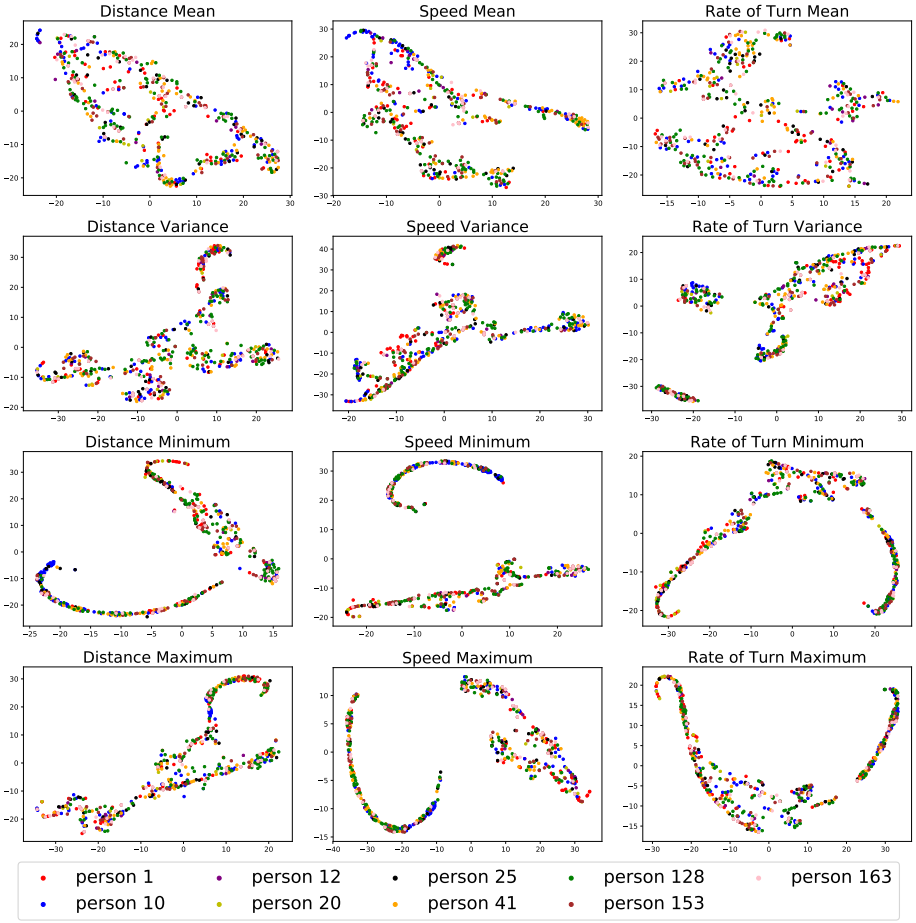


Fig. 1: tSNE representation of $\mathbf{MP_i}$. Each plot consists of 464 points representing trajectories from 9 different people. Clusters within the plots represent similar movement patterns between trajectories.

cover a very short distance after walking/cycling to complete the majority of the journey, is unusual.

The residual for each model was also computed to yield the residual probability density functions and in turn, the residual proximity matrices $\{\mathbf{RP_i}\}$. $\mathbf{R}$ and $\mathbf{M}$ were then calculated as the aggregated sum over all features for the residual and model proximity matrices respectively. In order to effectively combine the two matrices, $\mathbf{R}$ and $\mathbf{M}$ were normalised then combined using equation 16. A value of $\alpha = 0.1$ was experimentally selected to highlight the information available in the model residuals. The resulting matrix $\mathbf{D}$ was visualised using four different visualisation techniques to validate the distribution of the projected data.

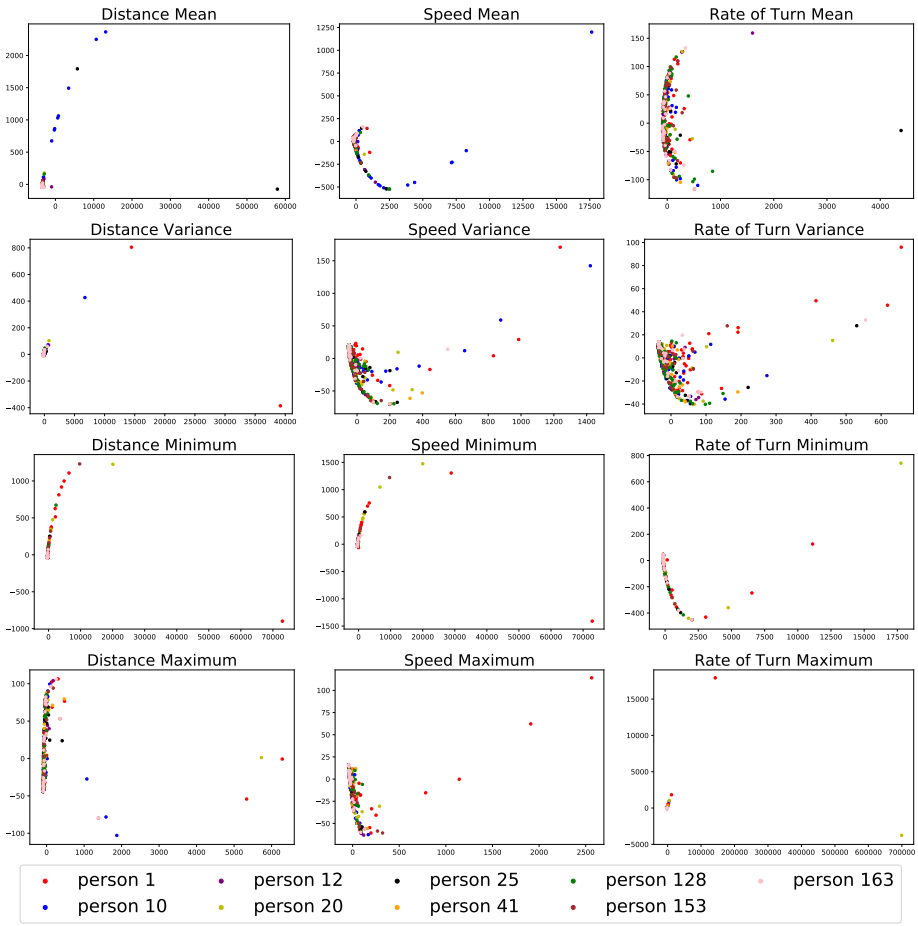


Fig. 2: PCA representation of $\mathbf{MP_i}$. Each plot consists of 464 points representing trajectories from 9 different people. Outliers represent anomalous trajectories.

Figure 7 shows the visualisation of $\mathbf{D}$. In order to achieve fully connected graphs for the LLE and Isomaps representations, the number of neighbours was experimentally set to 30. Although there are no obvious clusters, trajectories representing the same person are in close proximity, and tend to group. This is especially apparent when examining trajectories from persons 41 (orange) and 128 (green). It may also be the case that these groupings are separated more distinctly in higher dimensional space. Such patterns are meaningful since a single person's journeys often have intrinsic similarities and will be expected to group together when compared. This information was not found when using model analysis exclusively, however it is obvious with the inclusion of residual measures. Therefore, it can be concluded that model residuals contain information about the data that will otherwise be lost should residual analysis be neglected, as is the norm in literature.

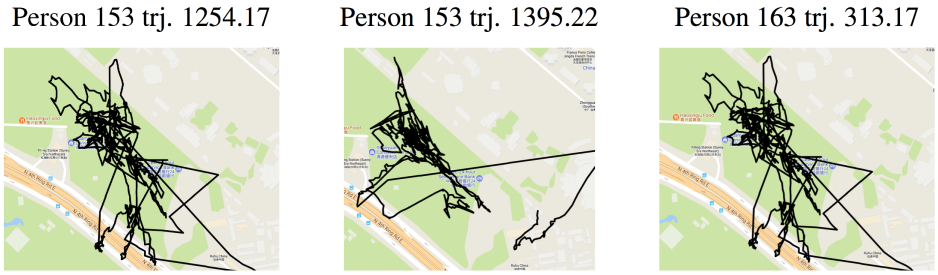


Fig. 3: Trajectories belonging to one cluster in the 'Speed Maximum' plot. The trips appear similar in nature and mostly comprise of walking movements.

Fig. 4: Trajectories belonging to one cluster in the 'Speed Variance' plot. The images depict journeys of travel on highways and main roads.

Fig. 5: Trajectories belonging to one cluster in the 'Speed Maximum' plot. The cluster mostly has trajectories of vehicular travel within main cities in China.

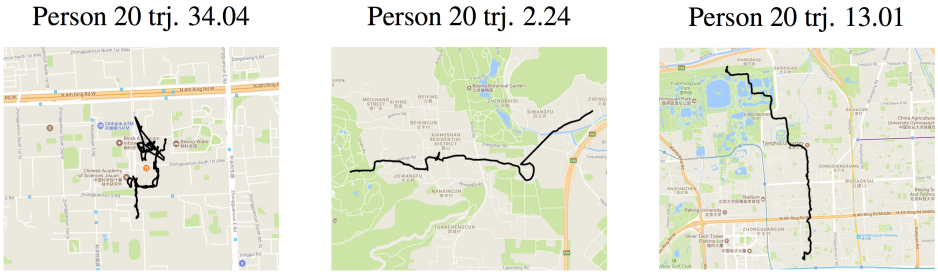


Fig. 6: Outlier trajectories in the 'Speed Variance' PCA plot. These trips differ from major clusters in the graph making them anomalous trajectories.

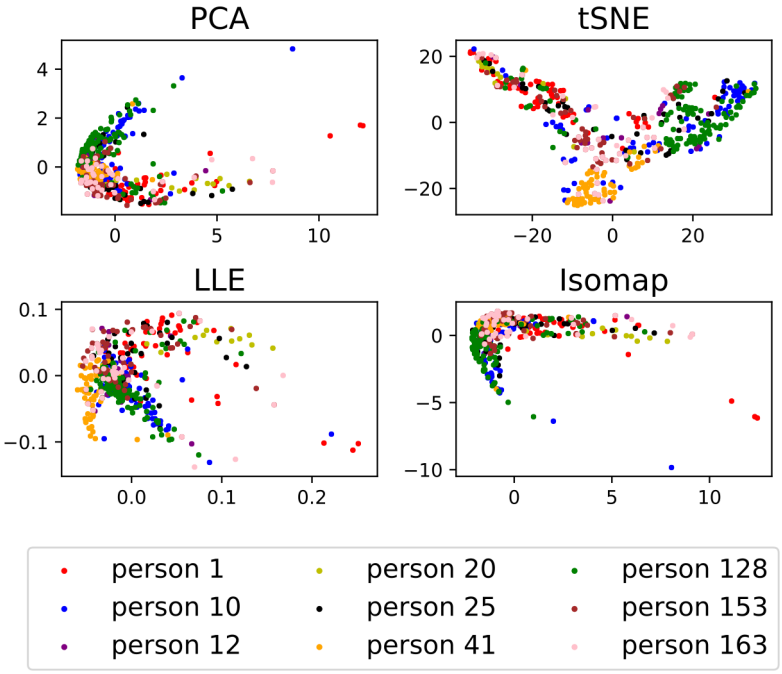


Fig. 7: PCA, tSNE, LLE and Isomap visualisations of the **D** proximity matrix. It can be seen that trajectories representing the same person are in close proximity, and tend to group

5 Conclusion

This paper has demonstrated an approach to detecting anomalous patterns of behaviour from GPS data may be achieved by a dissimilarity approach combining deterministic and stochastic time series models. It was possible to identify clusters relating to mobility patterns such as travelling between cities, driving in congested streets, or travelling by foot. Moreover, analysis of outliers visualisations has shown successful detection of anomalous journeys. The paper has also illustrated that information lost due to imperfect model fitting can be retrieved by residual analysis. By comparing between model residuals, each person's trajectories were grouped in close proximity; demonstrating the similarities between the movement patterns of a person across different trajectories.

Acknowledgements. Special thanks goes to Dr. Peter Van Manen and Living PlanIT AG for encouraging, supporting and funding this research.

References

1. Ramin Mehran, Alexis Oyama, and Mubarak Shah.: Abnormal crowd behavior detection using social force model. In: Computer Vision and Pattern Recognition (CVPR), 2009 IEEE Conference, pp. 935–942 (2009).
2. Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos.: Anomaly detection in crowded scenes. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference, pp. 1975–1981 (2010).
3. Dong-Her Shih, Ming-Hung Shih, David C Yen, and Jia-Hung Hsu.: Personal mobility pattern mining and anomaly detection in the GPS era. Cartography and Geographic Information Science, 43(1):55–67 (2016).
4. Naohiko Suzuki, Kosuke Hirasawa, Kenichi Tanaka, Yoshinori Kobayashi, Yoichi Sato, and Yozo Fujino. Learning motion patterns and anomaly detection by human trajectory analysis. In: Systems, Man and Cybernetics, 2007, ISIC. IEEE International Conference on, pp. 498–503. IEEE (2007).
5. Shan Jiang, Joseph Ferreira and Marta Gonza lez.: Clustering daily patterns of human activities in the city. Data Mining and Knowledge Discovery, 25(3):478510 (2012).
6. Ian Rice and David Lowe.: A decision support system to ease operator overload in multibeam passive sonar. IEEE Journal of Oceanic Engineering, (99):1–11 (2018).
7. Yu Zheng, Xing Xie, and Wei-Ying Ma.: Geolife: A collaborative social networking service among user, location and trajectory. IEEE Data Eng. Bull., 33(2):32–39 (2010).
8. Zhouyu Fu, Weiming Hu, and Tieniu Tan.: Similarity based vehicle trajectory clustering and anomaly detection. In: Image Processing, 2005. ICIP 2005. IEEE International Conference on, volume 2, pp. II–602. IEEE (2005).
9. Claudio Piciarelli, Christian Micheloni, and Gian Luca Foresti.: Trajectory-based anomalous event detection. IEEE Transactions on Circuits and Systems for video Technology, 18(11):1544–1554 (2008).
10. Glen Van Brummelen.: Heavenly mathematics: The forgotten art of spherical trigonometry, pp. 160–162. Princeton University Press (2012).
11. Richard J Martin.: A metric for ARMA processes. IEEE transactions on Signal Processing, 48(4):1164–1170 (2000).
12. Konstantinos Kalpakis, Dhiral Gada, and Vasundhara Puttagunta.: Distance measures for effective clustering of ARIMA time-series. In: Data Mining, 2001, ICDM 2001, Proceedings IEEE International Conference on, pages 273–280 (2001).
13. Laurens van der Maaten and Geoffrey Hinton.: Visualizing data using t-SNE. Journal of Machine Learning Research, 9(Nov):2579–2605, (2008)
14. Sam T Roweis and Lawrence K Saul.: Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500):2323–2326 (2000).
15. Joshua B Tenenbaum, Vin De Silva, and John C Langford.: A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500):2319–2323 (2000).

Model-based Data Exploration

Hans-Ulrich Kobialka, Daniel Paurat, Lisa Schrader

Fraunhofer Institute Intelligent Analysis and Information Systems IAIS
Fraunhofer Center for Machine Learning
Schloss Birlinghoven, 53757 Sankt Augustin, Germany
hans-ulrich.kobialka@iais.fraunhofer.de

Abstract. Data exploration is an approach of visually exploring data in order to understand the characteristics of the dataset. As both size and complexity of datasets increase substantially, data scientists take less look at the data directly but conduct experiments by training models and assess the outcome when applying these models on test data. We denote the use of ML models to experimentally obtain insights into the data at hand as *model-based data exploration* and show some examples from a recent industrial project.

Keywords: Root Cause Analysis, Failure Prediction, Offset Printing Machines, Data Quality, Data Labeling

1 Introduction

There are many publications on improving Machine Learning (ML) methods and demonstrating their achievements on benchmark data. These efforts have a strong focus on model optimization while leaving the benchmark data as is.

In contrast, in industrial projects, the sensor data available has usually never been used for ML purposes before. Issues of data quality are mainly unknown at the beginning of the project. Typically data quality issues become aware once models are built using ML methods and evaluation shows strange effects, e.g., on some data samples which are then inspected and discussed with domain experts in detail.

Industrial sensor data is high dimensional (i.e. many sensors), contains high frequency time series data, and may be created under many modes of operation. It is hard to plot in a comprehensive way and hard to interpret.

Therefore, inspection is a difficult and time consuming procedure, which can only be performed on few data samples. Interactive labeling by domain experts, as proposed by [11] e.g. on image segmentation tasks, is not a feasible approach for industrial senor data, because high-dimensional time series data is not the right medium to communicate to domain experts. Finding potentially interesting subspaces in the input data, which could simplify the communication to domain experts, is difficult when investigating statistics or distance measures in the input space because of its dimensionality and its time dynamics.

adfa, p. 1, 2011.

Our approach to this setting not to investigate the input data directly, but to conduct sequences of experiments by training ML models and assessing their outcome. Experiments are motivated by hypotheses (e.g. by information coming from domain experts). The outcome of experiments may or may not support these hypotheses, or may show interesting phenomena, which may give rise to new hypotheses/experiments. During this process, we obtain insights on the data and the learning problem. We denote the use of ML models, while conducting experiments on the data, as *model-based data exploration*.

The most obvious motivation to do model-based data exploration is to detect and filter low-quality data. This includes data containing measurement errors, data from non-calibrated sensors, data obtained under irregular operating conditions (e.g. machines runs in some cleaning mode, or some machine parts are not working as usual), or there is interference with other events (previous failures or external intervention). Outlier detection has been explored since decades and many ML methods, starting from simple clustering and regression methods to sophisticated ones, have been used for this purpose [2, 6].

Another well-known example of using ML methods for data exploration is feature selection [4] (e.g. backward deletion [9]) and sensitivity analysis [12].

When carrying this idea along to time series data, relevant events may happen in certain time periods only, leaving other parts of the time series being useless for the prediction task at hand.

Of course, most ML methods can cope with useless data if it covers only a small fraction of the total data. In that case, useless data can be regarded as noise. However, if that fraction becomes dominant, measures have to be taken to filter useless data, i.e. to improve signal-to-noise ratio.

Another motivation for model-based data exploration is building models on only a subset of the data. Suppose, for example, operating conditions and/or data collection differ between machines due to different software versions. In this case, it is worth building models for each machine and compare their performance with the one of a model trained on the data of all machines.

The bottom line of all these examples is that (lots of) experiments are conducted by building models on subsets of data and comparing their results. The focus is not on the quality of the particular model but the question of the experiment, e.g., can we regard some particular data samples as outliers, or not.

In this paper, we demonstrate procedures for model-based data exploration applied to an industrial project.

2 Failure Prediction for Printing Machines

Offset printing machines are big high-performance machines containing several printing units. On each printing unit, printing plates are mounted/replaced automatically for each page to be printed. This is a highly complex mechanical procedure. Failures

occurring during this procedure may have many causes. Knowing the reason of a particular failure and accumulation of such information should give valuable hints for maintenance, see www.sake-projekt.de.

The root cause of a particular failure may be some event seconds or even minutes before the failure. In case of the mounting of printing plates, the purpose of prediction is not to trigger any action for preventing the failure or mitigating its effect. Instead, a prediction model should identify the root cause. Technically, there should be one prediction model (classifier) for each root cause and the ones, which respond above some threshold, are considered for failure analysis.

Unfortunately, there is no knowledge about how root cause events are represented in the sensor data. We first extracted various features from the time series (thus transforming each multi-dimensional time series to a feature vector) and applied a bunch of well-known ML methods (Random Forest, Gradient Boosting Tree, Logistic Regression) with little success. Therefore, it can be assumed that, for the current application (plate change failure detection at printing machines), root cause events can't be detected by simple rules.

Furthermore, it is unknown which event caused a particular failure. It is assumed that these events are quite different in terms of the sensors involved, the time distance to failure, etc. Therefore, measures (e.g. in data pre-processing) may be beneficial for detecting some events, but not for others. Such effects may only be revealed when applied on particular subsets/clusters of the data. We were further told that some rare root causes are not detectable by the current set of sensors.

Besides this, there are data quality issues well known from similar projects: There may be data missing (time gaps in logging data). Data is collected from different sources having non-synchronized clocks. Sensor data comes from different machines, which are operated differently and may differ regarding software, sensors, material used etc.

2.1 Data Pre-Processing

The log channel, where the failure event is recorded, and the sensor data have different clocks, which show differences of one minute and more. For each failure event, we search in the sensor data within a time window of plus/minus 60 seconds for hints, which signal the occurrence of the failure. If no hint can be found, the failure is not considered. In case of ambiguity (multiple hints), decision is made under uncertainty.

Sensor data may contain gaps, i.e. timestamp values of consecutive data entries differ substantially. In probably most cases, these are artifacts, i.e. there is no sensor data skipped, but examples for information loss have been detected during data inspection. We filter time series containing time gaps of more than 10 seconds.

There is data of more than hundred sensors available. Domain experts selected the ones most relevant for the task, yielding 34 sensor channels. Data is collected at 50 Hz. Sensors range from being very high-frequent to changing only a few times throughout the time series. Domain experts told us to consider a time period of 5 minutes before the failure event. Therefore, the data for a particular failure event consists of a 34-dimensional time series of length 15,000 (i.e. 300 seconds, 50 Hz). There were data of

4,707 failures[1] available. We also collected 8,130 time series of periods where no failure occurred.

We regard time series of length 15,000 to be rather long, even as input to state-of-the-art deep neural networks [1]. We plotted some data and concluded that, even for high-frequent sensors, it is sufficient to segment the time series into non-overlapping intervals and to compute features of these intervals. Intervals have different length depending on their distance to the failure event, i.e. an interval, which is close in time to the failure event, is shorter than one containing sensor information recorded e.g. 4 minutes before the failure. This way we got 160 intervals, i.e. a time series of length 160.

For each interval, we computed some features, simply mean and standard deviation. This resulted into time series having length 160 and 68 dimensions (2 features for each of the 34 sensors).

Data normalization is done for each sensor separately, first removing the median and scaling the data according to the interquantile range (IQR). The IQR is the range between the 1st quartile and the 3rd quartile. This scales most values well but few can become quite large. Therefore, a non-linear scaling (tanh) is applied thereafter. Most sensors are normalized globally, i.e. using the data of all machines. A few sensors are not calibrated across all machines; for those sensors, we compute the average per machine and subtract it from all sensor values.

It turned out that Recurrent Neural Network (RNN) models show quite good precision-recall results when applied on this data. However, it was argued that this was because the failure is detected in the sensor data, but not the root cause, which is assumed to occur seconds of even minutes before.

Therefore, we cut off the sensor data up to one second from the time series. This way, it becomes a forecast task having a forecast horizon of 1 second. We increased the forecast horizon during later experiments, see section 5.

2.2 RNN learning task

We designed failure prediction as a regression task. Time series, which lead to failures, was given label 1, and time series without failure was labeled 0. An RNN was trained to compute from each time series a value y. Classification is done using a threshold t. If $y > t$, the model classifies the time series as failure, else not.

By varying threshold t, one can obtain different precision-recall results, e.g. $t = 1$ will yield precision close to 1 but very low recall. Lowering t will lower precision too and increase recall. This way, also the receiver operating characteristic (ROC) curve can be computed.

In the sequel, we denote a time series as a single *sample*, because from the regression's point of view, a time series is only one data point.

[1] In similar projects, often less failure data is available. From the application's point of view, it is preferred that failures occur as less as possible. However, ML methods suffer predicting rare failures having little statistics.

2.3 First Result

On the forecast task, we trained both Long Short Term Memory (LSTM) [5] and Echo State Network (ESN) [7, 10] models. Both were able to show almost perfect results on training data, but not on test data. Regularization (noise injection on LSTM, and ridge regression for ESN) improved this, but still results on test data were not acceptable. Figure 1 shows the test results for samples recorded from 14 printing machines sorted along the x-axis. For each machine, the non-failure time samples are printed first, and then followed by the ones for failure samples. For example the samples of the first machine are plotted at indices 0 to 95.

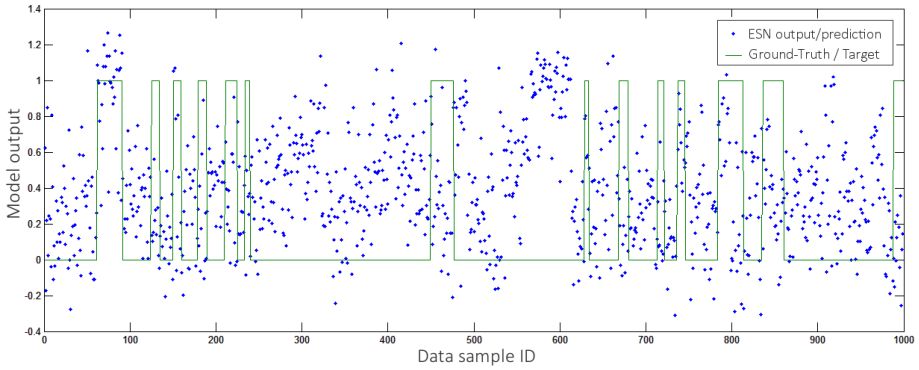


Fig. 1. ESN output (blue dots) for 1000 samples used for testing. The green line denotes the target labels (1 = failure, 0 = non-failure).

When faced with a disappointing result as in Figure 1, one can improve the model (e.g. using deep RNNs each layer having thousands of units; try out various regularization techniques, etc.). After some first try, we did not pursue this approach anymore, because we suspect low data quality.

We assume that there are problematic failure samples. This is because some types of failure cannot be detected by sensor data. For other failure samples, the corresponding sensor data may not correctly selected because of badly synchronized time stamps, and there may be other interfering events not recorded by the sensors selected.

3 Simplification of the Prediction Task

Domain experts reported to us that they do not expect a model to predict all failures, but if a failure is predicted, there should be much confidence (i.e. low false alarm rate)[2]. In other words, precision is first and recall is of minor priority. Therefore, we sought for models having high precision.

[2] We encounter this requirement in several other projects too.

We applied outlier detection on failure samples only. We followed the approach of Torr and Murray [13] by iteratively removing samples having biggest error and then re-train the model on the reduced data set.

Note that it is important that only failure samples are removed during this process, but *not* non-failure samples. Otherwise, even randomly labeled data samples could be perfectly separated by removing about half of the data samples.

During each iteration, we pruned 5% of failure samples having biggest train error. We also ensured that the number of failure samples per machine does not drop below 100[3]. We stopped after 10 iterations, so at the end 50% of all failure samples and 100% of all non-failure samples are left. In each iteration, we applied the current model to the whole data set and computed the mean prediction and standard deviation for failure samples and non-failure samples separately for the training set. Figure 2 shows that the margin between failure and non-failure samples is increasing during iteration. This improves the separability of the two classes and therefore validation results.

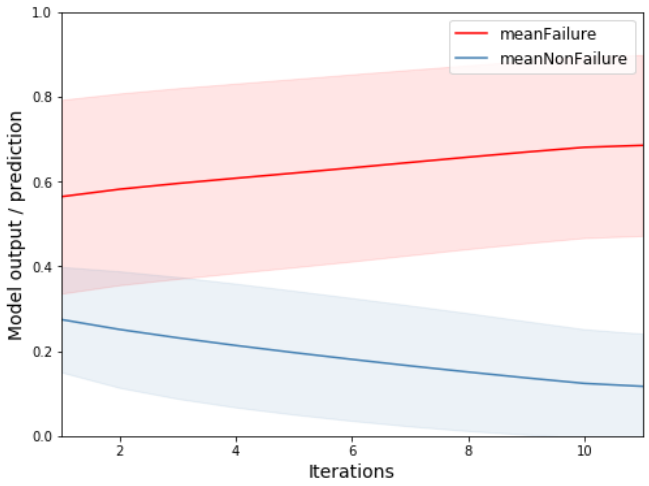


Fig. 2. Mean (+- standard deviation) of model output for failure samples and non-failure samples, first without removal of any sample, and then removing 5% failure samples after each step.

We applied 12-fold cross validation on the reduced data set. For different values for threshold *t*, precision and recall for the 12 test data sets are computed. Averaging test precision gives us an estimate about the precision to be expected when applying the model on unseen data samples. For the average test recall, it has to be considered that 50% of the failure samples are not included in this evaluation. So half of the average test recall can be expected for unseen data. So for threshold $t = 0.7$ (see table 1), the expected precision of the model is 0.95 and expected recall is 0.115.

[3] Otherwise, trained models tend to detect only failure samples of a few particular machines.

	Precision	Recall
Test fold 1	0.93	0.28
Test fold 2	1.00	0.21
Test fold 3	1.00	0.23
Test fold 4	0.94	0.22
Test fold 5	0.96	0.23
Test fold 6	0.88	0.18
Test fold 7	0.93	0.22
Test fold 8	0.98	0.25
Test fold 9	0.93	0.26
Test fold 10	0.98	0.21
Test fold 11	0.93	0.22
Test fold 12	0.97	0.18
Mean	0.95	0.23

Table 1. Test precision and recall for 12-fold cross validation, threshold $t = 0.7$.

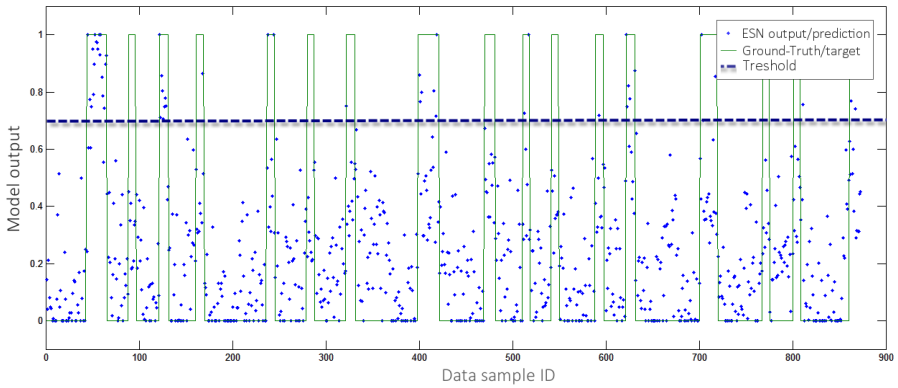


Fig. 3. Result on one of 12 test folds computed during cross validation. The line shows that for threshold $t = 0.7$, there is only one false positive, at position 811.

It is likely that pruning 50% of all failure samples filters not only outliers. We will refine this procedure later in order to improve our understanding of outliers and filter as less samples as possible. Note that at this point, the focus was *not* on perfect outlier detection. Instead, we wanted to get some model for testing design decisions, like in the following section.

As an alternative simplification of the learning task, we trained a model for each machine separately. Compared to a model trained on all data, none of the specialized models shows better performance. For some machines, in particular ones having only few training samples, the results were worse. Therefore, we did not pursue this approach any more.

4 Model-based assessment of preprocessing variants

In section 2.1, the various steps performed during preprocessing of the data are described. This included choices for parameters (e.g. interval length) which appear reasonable, but can be debated, of course. Once we have a prediction model, we can evaluate different parameter settings for preprocessing.

In particular, we were interested in the question, whether we should filter more samples during preprocessing, hopefully increasing data quality. Therefore, we experimented in

- narrowing the time window from plus/minus 60 seconds to plus/minus 30 seconds, and
- reducing the maximal tolerated gap between consecutive time stamps from 10 seconds to 2 seconds.

We tried out several choices to explore how many failure samples remain for training (Table 2). We see that, on the downside, filtering can reduce the amount of failure samples substantially.

	time window +/- 60 seconds	time window +/- 30 seconds
No filtering because of gaps	7,997	6,287
Gaps < 10 secs tolerated	**4,707**	
Gaps < 5 secs tolerated		3,605
Gaps < 2 secs tolerated	4,362	**3,337**
Gaps < 1 secs tolerated	2,611	

Table 2. The amount of failure samples left under different filtering options.

We compared two settings +/-60sec_gaps<10 and +/-30sec_gaps<2. As shown in the previous section, cross validation yielded an expected precision of 0.95 and an expected recall of 0.115 for +/-60sec_gaps<10. On the data set produced by +/-30sec_gaps<2, an expected precision of 0.94 and an expected recall of 0.082 is computed by cross validation.

Finally, we concluded, that a substantial reduction of training data can only be justified by a significant improvement of the ROC curve. This was clearly not the case, and so we continued with +/-60sec_gaps<10.

5 Exploration of forecast horizons

If the effect of a failure can be mitigated, it is desirable to predict its occurrence as early as possible. However, even if no mitigation is possible (as in the application at hand), the prediction time can give valuable hints to the root cause for the particular failure.

Failures may be caused by a broad spectrum of events, some of which may occur minutes before the failure; others may cause a failure directly. Knowing the earliest

time, when a failure can be reliably predicted, would help to group failures into failure classes. This would enable to train models specific for failure classes.

The previous sections describe experiments on data for forecast horizon of 1 second. By cutting off more time steps at the end of the time series (i.e. close to the failure event), we created data sets having forecast horizons of 4 seconds, 17 seconds and 79 seconds.

For later test, we selected the most recent 10% of the failure samples per machines. The remaining 90% of the failure samples were used for training prediction models for the above-mentioned forecast horizons according to the procedure introduced in section 3. The procedure was modified as follows: a) after each iteration, results (training samples used, error plots) are stored, b) iteration continues until almost no failure samples are left in the training data, and c) there is no longer a minimum number of failure samples per machine to be contained in the training data.

For each forecast horizon (1 second, 4 seconds, 17 seconds, 79 seconds), the results after each iteration are examined by manual inspection. First, we had a look after which iteration the margin between failure and non-failure samples has sufficiently large. Then we chose the result of one of the subsequent iteration steps, thereby determining the set of failure samples, which can be reliably predicted for the particular forecast horizon.

Again, we can adjust the trade-off between precision and the number of predictable failure samples. We used this degree of freedom to ensure that all failure samples, which are predictable 79 seconds before the failure event, are also contained in the set of failure samples, which can be predicted 17 seconds ahead. This is because, for each sample *x*, the time series used for 17-seconds prediction contain the time series used for 79-seconds prediction. This holds for shorter prediction horizons too.

$$x \in X_i \Rightarrow x \in X_j \qquad \text{for } i > j, \quad i, j \in \{1, 4, 17, 79\} \tag{1}$$

where X_i is the set of failure samples x which can be predicted for forecast horizon i. For each forecast horizon, we chose the iteration (i.e. the resulting failure sample set) such that overall constraint (1) is satisfied. This way, for forecast horizon i ($i \in \{1, 4, 17, 79\}$), X_i is identified for training model M_i.

We applied the models M_i on the failure samples of the test set. For 93% of the failure samples, on which one M_i computed an output > 0.6, constraint (1) was satisfied. These failure samples x can be classified having the largest prediction horizon i, for which $M_i(x) > 0.6$.

The failure samples, which were not in the test set, could be assessed (i.e. labeled) too in a cross validation scheme. In order to speed this up in the future, we are seeking for replacing the manual inspection step by some automatic procedure.

Furthermore, we will try to identify disjoint failure sets having root causes, which (according to the application experts) can be detected by disjoint sets of sensors. This will be performed in a similar way as above, just using another constraint ($X_i \cap X_j = \emptyset, i \neq j$).

6 Training of many models

Training of many models may be prohibitive depending on the method used and the computing sources available. The procedure described in section 3 requires the training of one model per iteration, plus a final 12-fold cross validation. This results in training of 22 models. [We also experimented using the cross validation error during each iteration[4]. With 12-fold cross validation, this gives 10 * 12 = 120 models trained.] We repeated this procedure many times, for example

- for refining the procedure (e.g. to ensure that failures of all machines are predicted),
- for collecting and refining metrics (RMSE, average values, standard deviations, F scores etc.)
- for assessment of preprocessing variants (section 4)
- for exploration of forecast horizons (section 5)
- for other outlier detection experiments

Up to now, we have trained several hundred models and we expect to train by far more than thousand models throughout the project.

We trained Echo State Networks (ESNs) of size 10,000. ESNs can be trained quite fast by training only the weights on the connections to the outputs using linear regression. In contrast, LSTMs or GRUs [3] train the weight matrices of the RNN and the connections between the input and the RNN as well.

Using a ESN of size 10,000, computation of the final RNN state for each of the 12,837 samples (i.e. time series of length 155) took about 5 hours on a laptop computer (Intel I7, 32 Giga Byte RAM). ESN training performs linear regression on these 12,837 states. This is done within 10 minutes. Training 6 models in parallel (during cross validation) takes about 30 minutes. Therefore, 12-fold cross validation takes 2 * 30 minutes.

ESN states can be re-used. Note that during cross validation and iterative pruning of outliers, different sets of training samples are used. Therefore, to train a model on a training set, the pre-computed states of these samples are collected, and then only linear regression has to be performed to train the output weight matrix. When re-using pre-existing states, the procedure of section 3 takes 10 * 10 minutes for the ten iterations, and 2 * 30 minutes for 12-fold cross validation. In total, this results in 2 hours and 40 minutes.

Using gradient descent, like with LSTM or GRU networks, states computed during previous training cannot be re-used, because the weight matrices, which are used to compute the state, are modified during training. This way, training takes much longer compared to ESNs, unless reducing network size by at least two magnitudes.

Using ESNs, we feel free to run as many experiments we want to. We think that this an important prerequisite to explore and validate ideas quickly.

[4] We abandoned this, because models tend not to detect failures of machines having only a few failure samples.

7 Conclusions

In this paper, we describe an experiment-driven working style which we call *model-based data exploration* and which we use in several industrial projects. Experiments are used, for example, to simplify a learning task (e.g. learning a model for each machine vs. learning a model on data of all machines), to filter outlier samples, to assess preprocessing variants, to do feature selection, and to classify failure samples by determining the time, when they can be predicted.

Interpretation of results is always an issue, For instance, once we classified samples according their forecast horizons, each class has to be investigated. We will determine important sensors for each class by feature selection & sensitivity analysis. In the end, findings have to be plotted and discussed with domain experts.

In all these experiments, neither the training data set nor the learning task is fixed. These are varied and evaluated by training models, and assessing their outcome. Optimization of hyper parameters of a model is not in the focus of these experiments.

Filtering of data (e.g. during preprocessing, exclusion of outlier samples, filtering data of other machines, feature selection, filtering not-needed time periods in time series data) can be viewed as improving the signal-to-noise ratio of the data.

Even the improvement of the signal-to-noise ratio may characterize the process, model-based data exploration has to be understood as gaining insights into the data at hand, and the application, which has generated the data. This process is centered on the data scientists who analyze data, generate hypotheses, set up experiments, and evaluate their outcome. This is some kind of data exploration, but this time, data scientists do not inspect the data directly but look at the outcome of ML models applied on the data.

Acknowledgement

This work was developed within the Fraunhofer Cluster of Excellence »Cognitive Internet Technologies« and partially supported by the BMWi project SAKE (Grant No. 01MD15006A).

References

1. Alom, Z., Taha, T., Yakopcic, C., Westberg, S., Hasan, M., Van Esesn, B., Awwal, A., Asari, V. (2018). The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. arXiv:1803.01164
2. Chandola, V., Banerjee, A., Kumar, V. (2009). Anomaly Detection: A Survey. ACM Computing Surveys 41(3), July 2009
3. Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 (2014).
4. Guyon, I., Elissee, A.: An introduction to variable and feature selection. Journal of Machine Learning Research, 3:1157-1182 (2003)
5. Hochreiter, S. and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, Vol. 9, No. 8, 1997, pp. 1735-1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735

6. Hodge, V. and Austin, J. 2004. A survey of outlier detection methodologies. Artificial Intelligence Review 22, 2, 85-126.
7. Jaeger, H., Haas, H.: Harnessing Nonlinearity: *Predicting Chaotic Systems and Saving Energy in Wireless Communication*. Science, April 2, 2004:78-80.
8. Kobialka H-U., Kayani U.: Echo State Networks with Sparse Output Connections. Proc. ICANN 2010, Lecture Notes in Computer Science 6352, pp.356~361, 2010.
9. Kohavi, R., John, G.: Wrappers for feature subset selection. Artificial Intelligence, 97 (1-2), 273-324 (1996)
10. Lukosevicius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. Computer Science Review, 3(3):127-149 (2009)
11. Nguyen, P., Ramanan, D., Fowlkes, C.: Active Testing: An Efficient and Robust Framework for Estimating Accuracy, arXiv:1807.00493, accepted to ICML 2018.
12. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. (2008). Global Sensitivity Analysis: The Primer. John Wiley & Sons.
13. Torr, P. and Murray, D. 1993. Outlier detection and motion segmentation. In Proceedings of SPIE, Sensor Fusion VI, Paul S. Schenker; Ed. Vol. 2059. 432-443.

Identification of nonstationary processes using noncausal bidirectional lattice filtering

Maciej Niedźwiecki and Damian Chojnacki*

Faculty of Electronics, Telecommunications and Computer Science,
Department of Automatic Control
Gdańsk University of Technology, ul. Narutowicza 11/12, Gdańsk, Poland
maciekn@eti.pg.edu.pl,damian.chojnacki@pg.edu.pl

Abstract. The problem of off-line identification of a nonstationary autoregressive process with a time-varying order and a time-varying degree of nonstationarity is considered and solved using the parallel estimation approach. The proposed parallel estimation scheme is made up of several bidirectional (noncausal) exponentially weighted lattice algorithms with different estimation memory and order settings. It is shown that optimization of both settings can be carried out by means of minimization of the locally evaluated accumulated forward/backward prediction error statistic.

Keywords: Identification of nonstationary processes, selection of model order, selection of estimation memory

1 Introduction

Autoregressive analysis is a popular modeling tool, used to solve practical problems in many different areas, such as biomedicine [1]–[3], geophysics [4]–[6], telecommunications [7]–[8] etc. When the analyzed processes are nonstationary, identification of their autoregressive models can be carried out using local estimation techniques, such as the well-known sliding-window or exponentially weighted least squares (EWLS) approaches. Local estimation algorithms are often called finite-memory since they relay on the limited (or effectively limited) number of signal samples. Owing to this property they are capable of tracking the time-varying signal parameters.

Two important decisions that must be taken when identifying the time-varying autoregressive model are the choice of the number of estimated autoregressive coefficients, i.e., the model order, and selection of the size of the local analysis interval, i.e., the estimation memory. Both decisions may have important quantitative (estimation accuracy) and qualitative (estimation adequacy) implications.

* This work was partially supported by the National Science Center under the agreement UMO-2015/17/B/ST7/03772. Calculations were carried out at the Academic Computer Centre in Gdańsk.

In this paper we will focus on noncausal estimation techniques, which can be applied when the analyzed signal is prerecorded and can be analyzed off-line. Noncausality means that at any given time instant t the local parameter estimates can be based on both "past" observations (collected prior to t) and "future" observations (collected after t). When applied to identification of nonstationary processes, noncausal estimators can significantly reduce the estimation bias (due to elimination of the so-called estimation delay, typical of all causal algorithms [9]).

In the proposed approach, which is a nontrivial modification of the method described in [10], noncausal estimates are obtained by combining results yielded by the exponentially weighted least squares lattice/ladder algorithms [11] running forward and backward in time, respectively. The problem of model order and estimation memory adaptation is solved using the parallel estimation approach. In this approach several competing algorithms, with different order and memory settings, are operated simultaneously and compared according to their locally assessed predictive abilities.

The proposed technique is computationally attractive and yields time-varying models with guaranteed uniform stability property which is important is such applications as parametric spectrum estimation.

2 Nonstationary autoregressive processes

Suppose that the analyzed discrete-time signal $\{y(t)\}$, $t = \ldots, -1, 0, 1, \ldots$, can be described or at least approximated by the following time-varying autoregressive (AR) model

$$\begin{gathered} y(t) = \sum_{i=1}^{n} a_{i,n}(t) y(t-i) + e_n(t) = \boldsymbol{\varphi}_n^{\mathrm{T}}(t) \boldsymbol{\alpha}_n(t) + e_n(t) \\ \mathrm{var}[e_n(t)] = \rho_n(t) \end{gathered} \tag{1}$$

where $\boldsymbol{\varphi}_n(t) = [y(t-1), \ldots, y(t-n)]^{\mathrm{T}}$ denotes regression vector, $\boldsymbol{\alpha}_n(t) = [a_{1,n}(t), \ldots, a_{n,n}(t)]^{\mathrm{T}}$ denotes the vector of autoregressive coefficients, and $\{e_n(t)\}$ denotes white noise with a time-dependent variance $\rho_n(t)$. In the sequel we will assume that the entire history of the signal $\{y(t), t = 1, \ldots, T_0\}$ is available, along with the "boundary" conditions $\{y(1-i), y(T_0+i), i = 1, \ldots, N\}$, where N denotes the maximum model order that will be considered.

When the driving noise variance $\rho_n(t)$ is bounded, $\boldsymbol{\alpha}_n(t)$ is a "sampled" version of a sufficiently smooth continuous time parameter trajectory, and at all time instants t all zeros of the characteristic polynomial $A[z, \boldsymbol{\alpha}_n(t)] = 1 - \sum_{i=1}^{n} a_{i,n}(t) z^{-i}$ are uniformly bounded away from the unit circle in the complex plane, the process (1) is uniformly exponentially stable [12]. According to the theory developed by Dahlhaus [13], under the conditions specified above $\{y(t)\}$ belongs to the class of locally stationary processes with uniquely defined

instantaneous spectral density function given by

$$S_n(\omega, t) = \frac{\rho_n(t)}{|A[e^{j\omega}, \boldsymbol{\alpha}_n(t)]|^2} \tag{2}$$

where $j = \sqrt{-1}$ and $\omega \in (-\pi, \pi]$ denotes the normalized angular frequency.

3 Equivalent parametrizations of a stationary autoregressive process

It is known that a zero-mean stationary AR process characterized by the set $\mathcal{P}_n = \{\rho_n, a_{1,n}, \ldots, a_{n,n}\}$ (further referred to as direct parametrization) can be equivalently specified in terms of autocorrelation coefficients $\mathcal{R}_n = \{r_0, r_1, \ldots, r_n\}$ where $r_i = \mathrm{E}[y(t)y(t-i)]$ (autocorrelation parametrization), or in terms of partial autocorrelation coefficients $\mathcal{Q}_n = \{r_0, q_1, \ldots, q_n\}$ where q_i is the normalized autocorrelation between $y(t)$ and $y(t-i)$ with the linear dependence on the intermediate variables $y(s), t-i < s < t$ removed (lattice parametrization).

All three parametrizations are equivalent, i.e., given any of them, one can determine the remaining two using invertible mappings

$$\begin{aligned}
\mathcal{P}_n &= F[\mathcal{R}_n], \quad \mathcal{R}_n = F^{-1}[\mathcal{P}_n] \\
\mathcal{R}_n &= G[\mathcal{Q}_n], \quad \mathcal{Q}_n = G^{-1}[\mathcal{R}_n] \\
\mathcal{Q}_n &= H[\mathcal{P}_n], \quad \mathcal{P}_n = H^{-1}[\mathcal{Q}_n].
\end{aligned}$$

Description of these mappings can be found e.g. in [14].

4 Causal lattice algorithm

The exponentially weighted least squares normalized lattice/ladder algorithm proposed by Lee, Morf and Friedlander [11], further referred to as EWLMF algorithm, is a time- and order-recursive estimation procedure known of its low computational cost and numerical robustness. The EWLMF algorithm is a lattice approximation of the EWLS algorithm. The EWLS algorithm, equipped with the forgetting constant λ_k, $0 < \lambda_k < 1$, provides a direct signal parametrization

$$\widehat{\mathcal{P}}_{n|k}(t) = \{\widehat{\rho}_{n|k}(t), \widehat{a}_{1,n|k}(t), \ldots, \widehat{a}_{n,n|k}(t)\}$$

where

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_{n|k}(t) &= [\widehat{a}_{1,n|k}(t), \ldots, \widehat{a}_{n,n|k}(t)]^{\mathrm{T}} \\
&= \arg\min_{\boldsymbol{\alpha}_n} \sum_{i=0}^{t-1} \lambda_k^i [y(t-i) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t-i)\boldsymbol{\alpha}_n]^2
\end{aligned} \tag{3}$$

$$\widehat{\rho}_{n|k}(t) = \frac{1}{L_k(t)} \sum_{i=0}^{t-1} \lambda_k^i [y(t-i) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t-i)\widehat{\boldsymbol{\alpha}}_{n|k}(t)]^2 \tag{4}$$

and $L_k(t) = \sum_{i=0}^{t-1} \lambda_k^i$ denotes the effective width of the applied exponential window. The explicit solution of (3) can be obtained in the form

$$\widehat{\boldsymbol{\alpha}}_{n|k}(t) = \widehat{\mathbf{R}}_{n|k}^{-1}(t)\widehat{\mathbf{r}}_{n|k}(t), \quad \widehat{\rho}_{n|k}(t) = \widehat{r}_{0|k}(t) - \widehat{\mathbf{r}}_{n|k}^{\mathrm{T}}(t)\widehat{\boldsymbol{\alpha}}_{n|k}(t) \tag{5}$$

where

$$\widehat{\mathbf{R}}_{n|k}(t) = \frac{1}{L_k(t)} \sum_{i=0}^{t-1} \lambda_k^i \boldsymbol{\varphi}_n(t-i)\boldsymbol{\varphi}_n^{\mathrm{T}}(t-i)$$

$$\widehat{\mathbf{r}}_{n|k}(t) = \frac{1}{L_k(t)} \sum_{i=0}^{t-1} \lambda_k^i y(t-i)\boldsymbol{\varphi}_n(t-i)$$

$$\widehat{r}_{0|k}(t) = \frac{1}{L_k(t)} \sum_{i=0}^{t-1} \lambda_k^i y^2(t-i) = \widetilde{r}_{0|k}(t).$$

The EWLMF algorithm estimates the normalized partial autocorrelation coefficients directly from the data, yielding the lattice signal parametrization

$$\widetilde{\mathcal{Q}}_{n|k}(t) = \{\widetilde{r}_{0|k}(t), \widetilde{q}_{1|k}(t), \ldots, \widetilde{q}_{n|k}(t)\}$$

The estimates $\widetilde{q}_{1|k}(t), \ldots, \widetilde{q}_{n|k}(t)$ are usually called reflection coefficients. Due to appropriate normalization, the estimates provided by the EWLMF algorithm obey the condition

$$|\widetilde{q}_{i|k}(t)| < 1, \quad \forall t, i = 1, \ldots, n \tag{6}$$

which guarantees that the corresponding AR models are at all times stable. Denote by

$$\widetilde{\mathcal{P}}_{n|k}(t) = H^{-1}[\widetilde{\mathcal{Q}}_{n|k}(t)] = \{\widetilde{\rho}_{n|k}(t), \widetilde{a}_{1,n|k}(t), \ldots, \widetilde{a}_{n,n|k}(t)\}$$

the direct parametrization that is an equivalent of the lattice parametrization yielded by the EWLMF algorithm. Since the EWLS algorithm does not guarantee model stability, it is clear that $\widehat{\mathcal{P}}_{n|k}(t) \neq \widetilde{\mathcal{P}}_{n|k}(t)$. We note, however, that both parametrizations become identical if the matrix $\widehat{\mathbf{R}}_{n|k}(t)$ and the vector $\widehat{\mathbf{r}}_{n|k}(t)$ appearing in (5) are replaced with

$$\widetilde{\mathbf{R}}_{n|k}(t) = \begin{bmatrix} \widetilde{r}_{0|k}(t) & & \widetilde{r}_{n-1|k}(t) \\ \vdots & \ddots & \vdots \\ \widetilde{r}_{n-1|k}(t) & & \widetilde{r}_{0|k}(t) \end{bmatrix}, \quad \widetilde{\mathbf{r}}_{n|k}(t) = \begin{bmatrix} \widetilde{r}_{1|k}(t) \ldots \widetilde{r}_{n|k}(t) \end{bmatrix}^{\mathrm{T}}$$

where

$$\widetilde{\mathcal{R}}_{n|k}(t) = \{\widetilde{r}_{0|k}(t), \widetilde{r}_{1|k}(t), \ldots, \widetilde{r}_{n|k}(t)\} = G[\widetilde{\mathcal{Q}}_{n|k}(t)]$$

denotes an autocorrelation parametrization equivalent to $\widetilde{\mathcal{Q}}_{n|k}(t)$. Therefore, the parametrization $\widetilde{\mathcal{P}}_{n|k}(t)$ can be regarded as a stable approximation of $\widehat{\mathcal{P}}_{n|k}(t)$.

5 Noncausal lattice algorithm

To obtain noncausal estimator of $\rho_n(t)$ and $\boldsymbol{\alpha}_n(t)$ we will combine results yielded by two lattice algorithms – the forward-time $(-)$ EWLMF algorithm equipped with a forgetting constant λ_{k^-}, providing the estimates

$$\widetilde{\mathcal{Q}}^{-}_{n|k}(t) = \{\widetilde{r}_{0|k^-}(t), \widetilde{q}_{1|k^-}(t), \ldots, \widetilde{q}_{n|k^-}(t)\}$$

and the backward-time $(+)$ EWLMF algorithm equipped with a forgetting constant λ_{k^+} providing the estimates

$$\widetilde{\mathcal{Q}}^{+}_{n|k}(t) = \{\widetilde{r}_{0|k^+}(t), \widetilde{q}_{1|k^+}(t), \ldots, \widetilde{q}_{n|k^+}(t)\}.$$

We will not assume that the forward and backward time EWLMF algorithms are equipped with the same forgetting constants. Setting $k^- \neq k^+$, one can fuse long-memory forward time estimation results with short-memory backward time ones or *vice versa*. Such asymmetric variants may be useful in the presence of abrupt parameter changes. Let $\pi = \{k^-, k^+\}$, $T_-(t) = \{1, \ldots, t-1\}$ and $T_+(t) = \{1, \ldots, T_0 - t\}$. The combined estimate can be obtained using a three-step procedure.

First, one can determine the autocorrelation parametrizations corresponding to $\widehat{\mathcal{Q}}^{-}_{n|k}(t-1)$ and $\widehat{\mathcal{Q}}^{+}_{n|k}(t+1)$

$$\widetilde{\mathcal{R}}^{\pm}_{n|k}(t \pm 1) = G[\widetilde{\mathcal{Q}}^{\pm}_{n|k}(t \pm 1)] = \{\widetilde{r}_{0|k^\pm}(t \pm 1), \widetilde{r}_{1|k^\pm}(t \pm 1), \ldots, \widetilde{r}_{n|k^\pm}(t \pm 1)\}$$

Since parametrizations $\widetilde{\mathcal{Q}}^{-}_{n|k}(t-1)$ and $\widetilde{\mathcal{Q}}^{+}_{n|k}(t+1)$ are stable, the covariance matrices made up of the estimates $\{\widetilde{r}_{i|k^-}(t), i = 0, \ldots, n\}$ and $\{\widetilde{r}_{i|k^+}(t), i = 0, \ldots, n\}$ must be positive definite [14]. Second, the two-sided autocorrelation parametrization

$$\widetilde{\mathcal{R}}_{n|\pi}(t) = \{\widetilde{r}_{0|\pi}(t), \widetilde{r}_{1|\pi}(t), \ldots, \widetilde{r}_{n|\pi}(t)\}$$

can be obtained using the formula

$$\widetilde{r}_{i|\pi}(t) = \mu_-(t)\widetilde{r}_{i|k^-}(t-1) + \mu_+(t)\widetilde{r}_{i|k^+}(t+1), \quad i = 0, \ldots, n \tag{7}$$

where $\mu_\pm(t) = L^{\pm}_{k^\pm}(t \pm 1)/L_\pi(t)$, $L_\pi(t) = L^{-}_{k^-}(t-1) + L^{+}_{k^+}(t+1)$ and $L^{\pm}_{k^\pm}(t \pm 1) = \sum_{i \in T_\pm(t)} \lambda^{i-1}_{k^\pm}$. Note that since the sequence $\{\widetilde{r}_{i|\pi}(t), i = 0, \ldots, n\}$ is a convex combination of $\{\widetilde{r}_{i|k^-}(t-1), i = 0, \ldots, n\}$ and $\{\widetilde{r}_{i|k^+}(t+1), i = 0, \ldots, n\}$, the parametrization $\widetilde{\mathcal{R}}_{n|\pi}(t)$ is at all times stable. Finally, based on $\widetilde{\mathcal{R}}_{n|\pi}(t)$, one can obtain the direct parametrization

$$\widetilde{\mathcal{P}}_{n|\pi}(t) = F[\widetilde{\mathcal{R}}_{n|\pi}(t)] = \{\widetilde{\rho}_{n|\pi}(t), \widetilde{a}_{1,n|\pi}(t), \ldots, \widetilde{a}_{n,n|\pi}(t)\}$$

The doubly exponentially weighted Lee-Morf-Friedlander (E^2WLMF) algorithm described above differs from the one proposed in [10] in one important aspect –

unlike [10] the obtained parameter estimates do not depend (in a deterministic sense) on the "central" sample $y(t)$.

Similarly as in the case of the EWLMF estimate, one can show that the E^{2}WLMF estimate $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t) = [\widetilde{a}_{1,n|\pi}(t), \ldots, \widetilde{a}_{n,n|\pi}(t)]^{\mathrm{T}}$ can be regarded as a "stable approximation" of the estimate obtained using the noncausal doubly exponentially weighted least squares (E^{2}WLS) algorithm

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{n|\pi}(t) &= [\widehat{a}_{1,n|\pi}(t), \ldots, \widehat{a}_{n,n|\pi}(t)]^{\mathrm{T}} \\ &= \arg\min_{\boldsymbol{\alpha}_n} \Big[\sum_{i=1}^{t-1} \lambda_{k^-}^{i-1} \{y(t-i) - [\boldsymbol{\varphi}_n^-(t-i)]^{\mathrm{T}} \boldsymbol{\alpha}_n\}^2 \\ &\qquad + \sum_{i=1}^{T_0-t} \lambda_{k^+}^{i-1} \{y(t+i) - [\boldsymbol{\varphi}_n^+(t+i)]^{\mathrm{T}} \boldsymbol{\alpha}_n\}^2 \Big]\end{aligned}$$

where $\boldsymbol{\varphi}_n^{\pm}(t) = [y(t \pm 1), \ldots, y(t \pm n)]^{\mathrm{T}}$. Actually, note that

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{n|\pi}(t) = &\Big[\mu_-(t)\widehat{\mathbf{R}}_{n|k^-}^-(t-1) + \mu_+(t)\widehat{\mathbf{R}}_{n|k^+}^+(t+1)\Big]^{-1} \\ &\times \Big[\mu_-(t)\widehat{\mathbf{r}}_{n|k^-}^-(t-1) + \mu_+(t)\widehat{\mathbf{r}}_{n|k^+}^+(t+1)\Big]\end{aligned} \tag{8}$$

where

$$\begin{aligned}\widehat{\mathbf{R}}_{n|k^\pm}^{\pm}(t \pm 1) &= \frac{1}{L_{k^\pm}^{\pm}(t \pm 1)} \sum_{i \in T_\pm(t)} \lambda_{k^\pm}^{i-1} \boldsymbol{\varphi}_n^{\pm}(t \pm i)[\boldsymbol{\varphi}_n^{\pm}(t \pm i)]^{\mathrm{T}} \\ \widehat{\mathbf{r}}_{n|k^\pm}^{\pm}(t \pm 1) &= \frac{1}{L_{k^\pm}^{\pm}(t \pm 1)} \sum_{i \in T_\pm(t)} \lambda_{k^\pm}^{i-1} y(t \pm i)\boldsymbol{\varphi}_n^{\pm}(t \pm i).\end{aligned}$$

Similarly, since $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t)$ must obey Yule-Walker equations defined in terms of $\{\widetilde{r}_{i|\pi}(t), i = 0, \ldots, n\}$ [14], it holds that

$$\begin{aligned}\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t) = &\Big[\mu_-(t)\widetilde{\mathbf{R}}_{n|k^-}^-(t-1) + \mu_+(t)\widetilde{\mathbf{R}}_{n|k^+}^+(t+1)\Big]^{-1} \\ &\times \Big[\mu_-(t)\widetilde{\mathbf{r}}_{n|k^-}^-(t-1) + \mu_+(t)\widetilde{\mathbf{r}}_{n|k^+}^+(t+1)\Big]\end{aligned}$$

where

$$\begin{aligned}\widetilde{\mathbf{R}}_{n|k^\pm}(t \pm 1) &= \begin{bmatrix} \widetilde{r}_{0|k^\pm}(t \pm 1) & \widetilde{r}_{n-1|k^\pm}(t \pm 1) \\ \vdots & \ddots & \vdots \\ \widetilde{r}_{n-1|k^\pm}(t \pm 1) & \widetilde{r}_{0|k^\pm}(t \pm 1) \end{bmatrix} \\ \widetilde{\mathbf{r}}_{n|k^\pm}(t \pm 1) &= \begin{bmatrix} \widetilde{r}_{1|k^\pm}(t \pm 1) & \ldots & \widetilde{r}_{n|k^\pm}(t \pm 1) \end{bmatrix}^{\mathrm{T}}.\end{aligned}$$

Hence, the estimates $\widehat{\boldsymbol{\alpha}}_{n|\pi}(t)$ and $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t)$ coincide if the quantities $\widehat{\mathbf{R}}_{n|k^\pm}^{\pm}(t \pm 1)$ and $\widehat{\mathbf{r}}_{n|k^\pm}^{\pm}(t \pm 1)$ are replaced in (8) with $\widetilde{\mathbf{R}}_{n|k^\pm}^{\pm}(t \pm 1)$ and $\widetilde{\mathbf{r}}_{n|k^\pm}^{\pm}(t \pm 1)$, respectively.

6 Model order and estimation memory adaptation

Based on $\widetilde{\mathcal{P}}_{n|\pi}(t)$, the parametric estimate of the instantaneous spectral density function $S_n(\omega, t)$ can be obtained in the form

$$\widetilde{S}_{n|\pi}(t) = \frac{\widetilde{\rho}_{n|\pi}(t)}{|A[e^{j\omega}, \widetilde{\boldsymbol{\alpha}}_{n|\pi}(t)]|^2} \tag{9}$$

where $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t) = [\widetilde{a}_{1,n|\pi}(t), \ldots, \widetilde{a}_{n,n|\pi}(t)]^{\mathrm{T}}$.

Selection of the order n of the autoregressive model, and the choice of forgetting factors $\lambda_{k^{\pm}}$ plays an important role in parametric spectral analysis. If the order is underestimated some important features of the resonant structure of $\{y(t)\}$ may be not revealed, while when it is overestimated some nonexistent resonances may be indicated. In both cases one may arrive at false qualitative conclusions. The optimal choice of λ_{k^-} and λ_{k^+}, i.e., the one that trades off the bias and variance components of the mean squared parameter estimation error, depends on the rate of parameter variation – forgetting factors should be smaller (which corresponds to shorter memory) when process parameters are subject to fast changes, and larger (which corresponds to longer memory) when parameters vary slowly with time.

Our solution to the order/memory optimization problem will be based on parallel estimation. Consider several E^{2}WLMF algorithms with different order and memory settings, running in parallel. Denote by $\mathcal{N} = \{1, \ldots, N\}$ the set of all model orders that will be considered, and by Π the set of all considered pairs $\pi = \{k_-, k_+\}$. The data-adaptive version of (9) can be expressed in the form

$$\widetilde{S}_{\widehat{n}(t)|\widehat{\pi}(t)}(t) = \frac{\widetilde{\rho}_{\widehat{n}(t)|\widehat{\pi}(t)}(t)}{|A[e^{j\omega}, \widetilde{\boldsymbol{\alpha}}_{\widehat{n}(t)|\widehat{\pi}(t)}(t)]|^2} \tag{10}$$

where

$$\{\widehat{n}(t), \widehat{\pi}(t)\} = \{\widehat{n}(t), \widehat{k}_-(t), \widehat{k}_+(t)\} = \arg \min_{\substack{n \in \mathcal{N} \\ k \in \mathcal{K}}} J_{n|\pi}(t)$$

and $J_{n|\pi}(t)$ denotes the local decision statistic.

The proposed selection criterion takes advantage of the fact that, unlike the estimates considered in [10], the estimates $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t)$ are not functions of $y(t)$ and therefore they can be used to compute unbiased forward and backward prediction errors

$$\varepsilon^{\pm}_{n|\pi}(t) = y(t) - [\boldsymbol{\varphi}^{\pm}_n(t \pm 1)]^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}}_{n|\pi}(t).$$

Consequently, one can adopt for $J_{n|\pi}(t)$ the following prediction error (PE) statistic

$$J_{n|\pi}(t) = \sum_{i=-M}^{M} [\boldsymbol{\varepsilon}^{-}_{n|\pi}(t-i)]^2 + \sum_{i=-M}^{M} [\boldsymbol{\varepsilon}^{+}_{n|\pi}(t+i)]^2 \tag{11}$$

where $M \in [20, 50]$ is the parameter that controls the size of the local decision window $[t - M, t + M]$ centered around t.

7 Computational complexity

Denote by $K_\pi \leq K(K+1)/2$ the number of considered forward-backward pairs $\pi = (k^-, k^+)$. For the assumed maximum model order N the per sample computational load (the number of multiply-add operations) of the proposed parallel estimation scheme is pretty low and is approximately equal to

$$l(N) = 2KA(N) + 2KB(N) + K_\pi C(N)$$

where $A(N) = 30N$ denotes the load of the ELMF algorithm (given that the Newton-Raphcon method is used to evaluate square roots), $B(N) = 2N + N^2$ denotes the load of the G transform (computation of autocorrelation coefficients based on reflection coefficients), and $C(N) = 2 + 4N + N^2$ is the load of the F transform (computation of autoregressive coefficients based on autocorrelation coefficients). Note that the first stage of processing is the computationally cheapest one and that the only quantities that have to be memorized during the forward/backward sweep of the EWLMF algorithms are the forward/backward reflection coefficients.

8 Simulation results

To verify the proposed order and estimation memory selection rule, a nonstationary variable-order autoregressive process was generated. Process generation was based on 4 time-invariant AR anchor models M_1, M_2, M_3 and M_4, of orders 2, 4, 6 and 8, respectively. The characteristic polynomial $A_i(z)$ of the model M_i had i pairs of complex-conjugate zeros, given by $z_k^\pm = 0.995e^{\pm jk\pi/5}, k = 1, \ldots, i$. The generated signal $\{y(t), t = 1, \ldots, T_0\}$ had stationary periods, during which it was governed by anchor models, and nonstationary periods, when the generating model was obtained by morphing one anchor model into another one. Transition from M_{i-1} to M_i was realized by moving, with a constant speed, the i-th pair of complex-conjugate zeros from their initial zero positions towards the unit circle – see Fig. 1. The simulation scenario is symbolically depicted in Fig. 1. Note that according to this scenario the order of the generating model gradually increased from 2 to 8.

The adopted value of T_0 was equal to 5000 and the breakpoints, marked with bullets in Fig. 1, had the following time coordinates: $t_1 = 1000$, $t_2 = 1500$, $t_3 = 2500$, $t_4 = 3000$, $t_5 = 4000$, $t_6 = 4500$. The parallel estimation scheme was made up of 4 E^2WLMF algorithms combining results yielded by $K = 3$ forward/backward EWLMF trackers equipped with forgetting constants $\lambda_1 = 0.95$, $\lambda_2 = 0.99$ and $\lambda_3 = 0.995$. The 4 combinations of forward/backward forgetting constants were: (0.99, 0.99), (0.995, 0.995), (0.995, 0.95) and (0.95, 0.995), which corresponds to $\pi_1 = (2, 2)$, $\pi_2 = (3, 3)$, $\pi_3 = (3, 1)$ and $\pi_4 = (1, 3)$, respectively. The parameter M which determines the width of the local decision window was set to 50.

Two measures of fit were used to evaluate identification results: the mean squared parameter tracking error and the Itakura-Saito spectral distortion measure (see Table 1), both averaged over $t \in [1, T_0]$ and 100 independent realizations

of $\{y(t)\}$. Table 1 compares results yielded by 3 unidirectional $(\lambda_1, \ldots, \lambda_3)$ and 4 bidirectional $(\pi_1, \ldots, \pi_4)$ lattice algorithms (for different values of n), with the results yielded by the proposed adaptive scheme (for different values of N). Note that when the model order is not underestimated $(n, N \geq 8)$ the algorithm with adaptive order and memory assignment (A) provides results that are uniformly the best, irrespective of the choice of N.

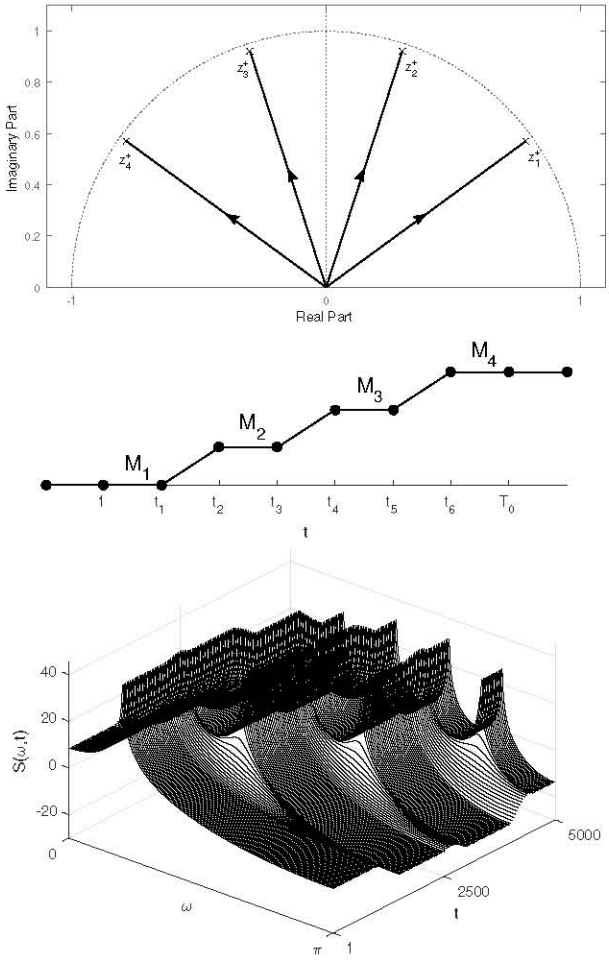


Fig. 1. Trajectories of zeros of the characteristic polynomial (top figure), the applied simulation scenario (middle figure), and the corresponding time-varying spectral density function (bottom plot).

Table 1. Averaged Itakura-Saito distortion measures (left table) and mean square parameter estimation errors (right table).

n/N	λ_1	λ_2	λ_3	π_1	π_2	π_3	π_4	A
1	4,600	4,266	4,199	4,185	4,131	4,193	4,170	4,155
2	3,183	2,751	2,796	2,551	2,552	2,697	2,644	2,603
3	3,093	2,619	2,660	2,398	2,397	2,559	2,488	2,446
4	2,092	1,536	1,616	1,318	1,358	1,482	1,483	1,357
5	2,169	1,536	1,611	1,298	1,333	1,472	1,464	1,338
6	1,118	0,577	0,711	0,452	0,586	0,563	0,726	0,437
7	1,180	0,583	0,697	0,416	0,519	0,558	0,629	0,415
8	0,775	0,144	0,208	0,070	0,163	0,126	0,239	**0,067**
9	0,848	0,147	0,187	0,071	0,146	0,117	0,206	**0,068**
10	0,925	0,154	0,189	0,072	0,134	0,120	0,192	**0,068**
11	1,006	0,160	0,191	0,073	0,125	0,122	0,183	**0,069**
12	1,093	0,167	0,194	0,075	0,121	0,124	0,180	**0,069**
13	1,187	0,174	0,196	0,077	0,120	0,126	0,182	**0,069**
14	1,301	0,181	0,198	0,077	0,116	0,129	0,179	**0,069**
15	1,413	0,190	0,204	0,080	0,114	0,132	0,178	**0,070**
16	1,547	0,198	0,208	0,082	0,112	0,135	0,178	**0,070**
17	1,674	0,206	0,213	0,084	0,111	0,138	0,179	**0,071**
18	1,821	0,214	0,218	0,086	0,112	0,142	0,182	**0,071**
19	1,952	0,221	0,222	0,088	0,111	0,145	0,183	**0,071**
20	2,095	0,231	0,227	0,091	0,111	0,148	0,185	**0,072**

n/N	λ_1	λ_2	λ_3	π_1	π_2	π_3	π_4	A
1	12,027	11,986	11,956	12,002	11,992	11,959	12,049	12,010
2	8,673	8,623	8,577	8,679	8,685	8,587	8,789	8,681
3	6,504	6,386	6,339	6,418	6,444	6,339	6,615	6,484
4	3,015	2,897	2,839	3,011	3,291	2,829	3,729	2,984
5	2,628	2,460	2,566	2,315	2,517	2,514	2,720	2,402
6	1,106	1,026	1,282	0,593	0,863	1,156	1,113	0,815
7	1,083	0,547	0,675	0,723	2,412	0,596	2,970	0,478
8	1,102	0,348	0,441	0,369	1,522	0,352	2,125	**0,236**
9	1,392	0,425	0,523	0,389	1,284	0,436	1,718	**0,252**
10	1,688	0,485	0,552	0,444	1,347	0,464	1,749	**0,266**
11	2,014	0,551	0,578	0,497	1,449	0,486	1,852	**0,278**
12	2,340	0,628	0,618	0,560	1,680	0,523	2,135	**0,294**
13	2,613	0,690	0,646	0,549	1,560	0,546	2,067	**0,298**
14	2,947	0,761	0,682	0,557	1,377	0,580	1,910	**0,305**
15	3,197	0,827	0,716	0,579	1,275	0,613	1,801	**0,312**
16	3,614	0,904	0,756	0,615	1,245	0,646	1,778	**0,322**
17	3,887	0,962	0,784	0,653	1,306	0,672	1,855	**0,328**
18	4,184	1,026	0,816	0,673	1,302	0,702	1,868	**0,333**
19	4,475	1,085	0,847	0,691	1,259	0,731	1,841	**0,339**
20	4,805	1,160	0,885	0,723	1,223	0,766	1,820	**0,345**

Our second example shows the result of application of the proposed approach to analysis of a real signal. Fig. 2 shows the plots of 5 fragments of a speech signal (sampled at the rate of 22.05 kHz) and the corresponding estimates of the time-varying spectrum obtained using the parallel estimation scheme described above (with the same settings).

9 Conclusion

A new noncausal (bidirectional) lattice filtering algorithm was designed for off-line identification of nonstationary autoregressive processes and an adaptive mechanism was proposed for dynamic selection of the number of estimated coefficients and the most appropriate estimation memory, matching the degree of process nonstationarity. It was shown that the proposed adaptive parallel estimation scheme outperforms the fixed-order fixed-memory algorithms it is made up of.

References

1. T. Wada, M. Jinnouchi and Y. Matsumura, "Application of autoregressive modelling for the analysis of clinical and other biological data," *Ann. Inst. Statist. Math.*, vol. 40, pp. 211-227, 1998.
2. V.K. Jirsa and A.R. McIntosh, Eds., *Handbook of Brain Connectivity*, Springer-Verlag, 2007.
3. R. Takalo, H. Hytti, H. Ihalainen and A. Sohlberg, *Adaptive autoregressive model for reduction of noise in SPECT*, Comp. Math. Methods in Medicine, vol. 2015, 9 p., 2015.
4. C. Li and R.L. Nowack, "Application of autoregressive extrapolation to seismic tomography," Bull. Seism. Soc. Amer., vol. 94, pp. 1456?-1466, 2004.

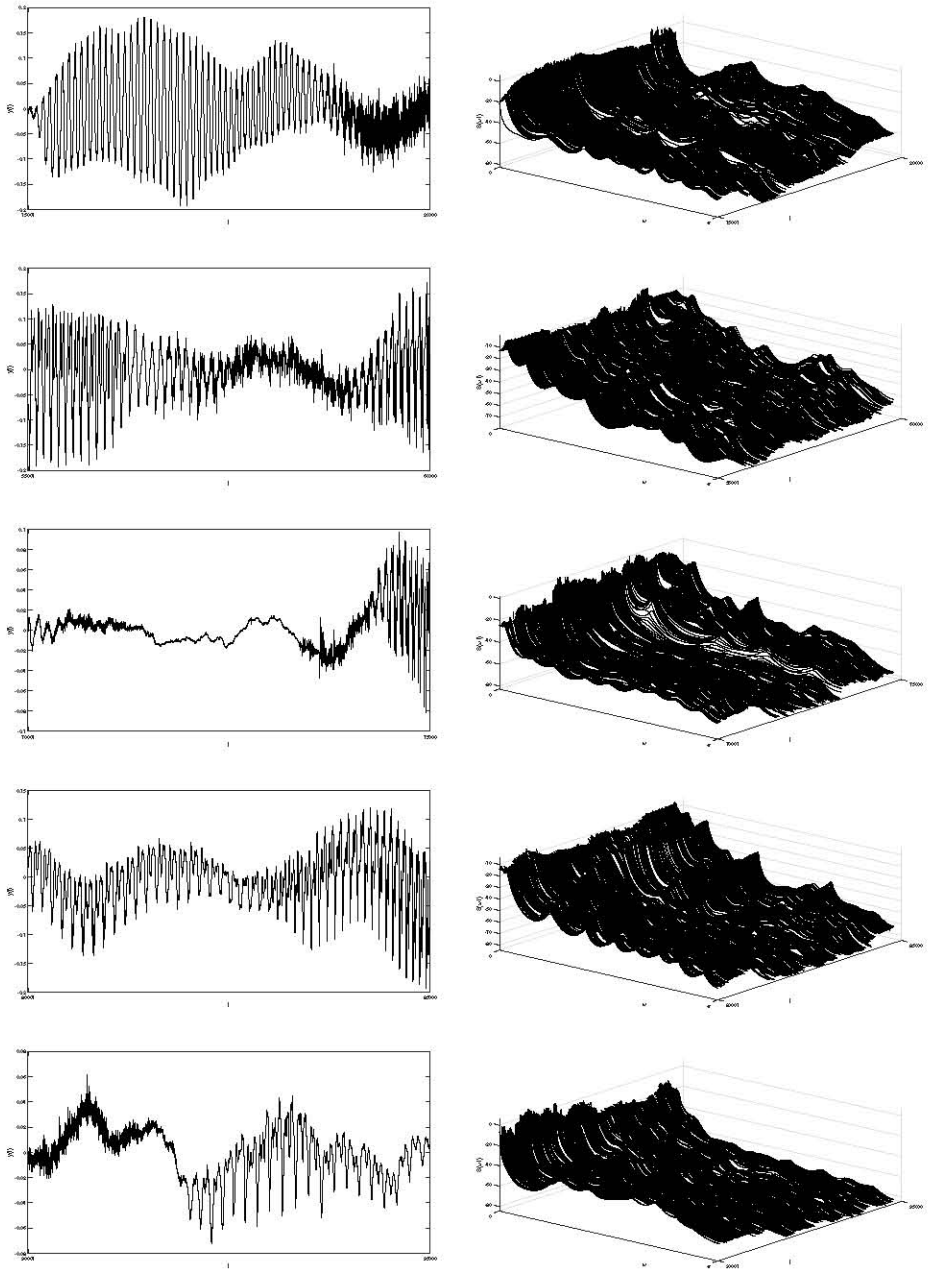


Fig. 2. Five fragments of a speech signal (left figures) and the estimated time-varying spectra (right figures).

5. P. Lesage, F. Glangeaud and J. Mars, "Applications of autoregressive models and time-frequency analysis to the study of volcanic tremor and long-period events," *J. Volc. Geotherm. Res.*, vol. 114, pp. 391?-417, 2002.
6. D. Brillinger, E.A. Robinson and F.P. Schoenberg, Eds., *Time Series Analysis and Applications to Geophysical Systems*, Springer, 2012.
7. K.E. Baddour, N.C. Beaulieu, "Autoregressive models for fading channel simulation," IEEE Trans. Wireless Comm., vol. 4, pp. 1650–1662, 2005.
8. J.F. Hayes and T.V.J. Ganesh Babu, *Modeling and Analysis of Telecommunication Networks,* Wiley, 2004.
9. M. Niedźwiecki, *Identification of Time-varying Processes*, Wiley, 2000.
10. M. Niedźwiecki, M. Meller, and D. Chojnacki, "Lattice filter based autoregressive spectrum estimation with joint model order and estimation bandwidth adaptation," *Proc. 56th IEEE Conference on Decision and Control*, Melbourne, Australia, pp. 4618–4625, 2017.
11. D.T.L. Lee, M. Morf, and B. Friedlander, "Recursive least squares ladder estimation algorithms," *IEEE Trans. Acoust., Speech, Signal Process.* , vol. 29, pp. 627–641, 1981.
12. E. Moulines, P. Priouret, and F. Roueff, "On recursive estimation for time-varying autoregressive processes," *Ann. Statist.*, vol. 33, pp. 2610–2654, 2005.
13. R. Dahlhaus, "Locally stationary processes," *Handbook Statist.*, vol. 25, pp. 1–37, 2012.
14. B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE* , vol. 70, pp. 829–867, 1982.

Likelihood based inference for an Identifiable Fractional Vector Error Correction Model

Federico Carlini*, Katarzyna Łasak†

6 September 2018

Abstract

We consider the Fractional Vector Error Correction model proposed in Avarucci (2007), which is characterized by a richer lag structure than the models proposed in Granger (1986) and Johansen (2008, 2009). In particular, we discuss the properties of the model of Avarucci (2007) (FECM) in comparison to the model of Johansen (2008, 2009) (FCVAR). Both models generate the same class of processes, but the properties of the two models are different. First, opposed to the model of Johansen (2008, 2009), the model of Avarucci has a convenient nesting structure, which allows for testing the number of lags and the cointegration rank exactly in the same way as in the standard I(1) cointegration framework of Johansen (1995) and hence might be attractive for econometric practice. Second, we find that the model of Avarucci (2007) is almost free from identification problems, contrary to the model of Johansen (2008, 2009) and Johansen and Nielsen (2012), which identification problems are discussed in Carlini and Santucci de Magistris (2017). However, due to a larger number of parameters, the estimation of the FECM model of Avarucci (2007) turns out to be more complicated. Therefore, we propose a 4-step estimation procedure for this model that is based on the switching algorithm employed in Carlini and Mosconi (2014), together with the GLS procedure of Mosconi and Paruolo (2014). We check the performance of the proposed estimation procedure in finite samples by means of

*USI, Lugano, E-mail: federico.carlini@usi.ch

†Corresponding author. Tinbergen Institute and University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, Email: k.a.lasak@uva.nl.

a Monte Carlo experiment and we prove the asymptotic distribution of the estimators of all the parameters. The solution of the model has been previously derived in Avarucci (2007), while testing for the rank has been discussed in Łasak and Velasco (for cointegration strength >0.5) and Avarucci and Velasco (for cointegration strength <0.5). Therefore our paper fills in the gap for a complete inference based on Avarucci (2007) model.

Keywords: Error correction model, Gaussian VAR model, Fractional Cointegration, Estimation algorithm, Maximum likelihood estimation, Switching Algorithm, Reduced Rank Regression. **JEL:** C13, C32.

1 Introduction

The econometrics literature on fractional co-integration has developed rapidly in recent years. An empirically attractive modeling strategy is to use parametric inference, based on an econometric model that fully describes the system under consideration. It allows identification of the long-run and short-run structure of the model, as well as of the common stochastic trends and the impulse response functions summarizing the system dynamics. Three different Fractional Vector Error Correction Models (FVECM) have been proposed in the literature due to Granger (1986), Johansen (2008, 2009) and Avarucci (2007). These models turn out to be almost identical in the simplest case without short run dynamics, but more generally they are characterized by different lag structure specifications.

The FCVAR model proposed in Johansen (2008, 2009) has a convenient algebraic structure. The inference for this model has been developed in Johansen and Nielsen (2012). However, there exist identification problems in this model, as mentioned in Johansen and Nielsen (2012) and further discussed in Carlini and Santucci de Magistris (2017).

In this paper we demonstrate that the FECM model proposed by Avarucci (2007) is almost free from identification problems contrary to the FCVAR model of Johansen (2008, 2009). Also, designing testing procedures for the lag length and the cointegration rank is straightforward in FECM, due to the fact that the nesting structure follows the usual structure known for the I(1)

Cointegrated Vector AutoRegressive (CVAR) model. However, the estimation is more complicated in thie FECM model due to the multiplicative structure of the parameters involved.

We conclude that the model proposed by Avarucci might be more convenient for practitioners if we design an estimation procedure for this model, which we do in this paper.

We propose a 4 step algorithm, which is based on the approach of Carlini and Mosconi (2014) that maximizes the profile likelihood function using a switching algorithm and implements the GLS procedure proposed in Mosconi and Paruolo (2014). We check the performance of the proposed estimation procedure in finite samples by means of a Monte Carlo experiment and we prove the asymptotic distribution of the estimators of all the parameters. The solution of the model has been previously derived in Avarucci (2007), while testing for the rank has been discussed in Łasak and Velasco (for cointegration strength >0.5) and Avarucci and Velasco (for cointegration strength <0.5). Therefore our paper fills in the gap for a complete inference based on Avarucci (2007) model.

The remainder of the paper is organized as follows. Section 2 presents the FCVAR model proposed in Johansen (2008, 2009) and the FECM model suggested in Avarucci (2007). In particular we discuss derivation and we give the solutions of both models. We also present different representations of FECM. In Section 3 we discuss the identification and the convenient nesting structure of the FECM model. Section 4 introduces the profile likelihood and a 4 step switching algorithm to estimate the parameters of the FECM model together with the asymptotic distribution of the estimators of all the parameters. Section 5 illustrates the small sample properties of our estimation procedure by means of a Monte Carlo experiment. Section 6 concludes. Appendix A describes how to estimate a bilinear form with a GLS model. Appendix B presents Figures and Tables of the Monte Carlo experiment.

2 Model comparison

2.1 Johansen's FCVAR model

The model of Johansen (2008, 2009), which we denote in this paper as FCVAR, is given by the following dynamics

$$\Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} \Gamma_j \Delta^d L_b^j X_t + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \Omega), \tag{1}$$

where the vector of variables X_t is $p-$dimensional, the loadings α and the cointegrating relations β are $p \times r$ matrices with $0 \leq r \leq p$, Γ_j are $p \times p$ matrices of the short run dynamics and the fractional difference operator is given by the binomial expansion $\Delta^d := (1-L)^d = \sum_{j=0}^{\infty} (-1)^j \binom{d}{j} L^j$ and the fractional lag operator is defined $L_b = 1 - \Delta^b$.

As shown in Johansen (2008, 2009), this model could be derived from the standard VAR model $\Delta Y_t = \alpha\beta' L Y_t + \sum_{j=1}^{k} \Gamma_j \Delta L^j Y_t + \varepsilon_t$ analysed in Johansen (1995), where the lag operator L is such that $LX_t = X_{t-1}$, and the difference operator $\Delta = 1 - L$, in the following way. First replace the difference operator Δ and the lag operator $L = 1 - \Delta$ by fractional difference operator Δ^b and the fractional lag operator $L_b = 1 - \Delta^b$, respectively, to obtain $\Delta^b Y_t = \alpha\beta'(1-\Delta^b) Y_t + \sum_{j=1}^{k} \Gamma_j \Delta^b L_b^j Y_t + \varepsilon_t$. Next define $Y_t = \Delta^{d-b} X_t$ to get the model (1).

The Granger representation of the model (1) is given in Johansen (2008, 2009):

$$X_t = C\Delta_+^{-d} \varepsilon_t + \Delta_+^{-(d-b)} Y_t^+ + \mu_t, \tag{2}$$

where μ_t is a deterministic component generated by initial values, $C = \beta_\perp (\alpha_\perp' \Gamma \beta_\perp)^{-1} \alpha_\perp'$ and $Y_t^+ = \sum_{n=0}^{t-1} \tau_n \varepsilon_{t-n}$, so Y_t^+ is fractional of order zero. Thus the solution of model (1) implies that X_t is a process integrated of order d ($I(d)$), while $\Delta^b X_t$ and $\beta' X_t$ are $I(d-b)$.

In general the cointegration rank r and number of lagged differences k is not known and needs to be determined. However, the nesting structure, as described in Carlini and Santucci di Magistris

(2017) turns out to be of the following form:

$$\begin{array}{ccccccccc}
\mathcal{H}_{0,0} & \subset & \mathcal{H}_{0,1} & \subset & \mathcal{H}_{0,2} & \subset & \dots & \subset & \mathcal{H}_{0,k} \\
\cap & & \cap & & \cap & & & & \cap \\
\mathcal{H}_{1,0} & \subset & \mathcal{H}_{1,1} & \subset & \mathcal{H}_{1,2} & \subset & \dots & \subset & \mathcal{H}_{1,k} \\
\cap & & \cap & & \cap & & & & \cap \\
\vdots & & \vdots & & \vdots & & \ddots & & \vdots \\
\cap & & \cap & & \cap & & & & \cap \\
\mathcal{H}_{p,0} & \subset & \mathcal{H}_{p,1} & \subset & \mathcal{H}_{p,2} & \subset & \dots & \subset & \mathcal{H}_{p,k}
\end{array}$$

with

$$\begin{array}{ccc}
\mathcal{H}_{0,1} & \subset & \mathcal{H}_{p,0} \\
\mathcal{H}_{0,2} & \subset & \mathcal{H}_{p,1} \\
\vdots & & \\
\vdots & & \\
\mathcal{H}_{0,k} & \subset & \mathcal{H}_{p,k-1}
\end{array},$$

where $\mathcal{H}_{p,k}$ denotes the hypothesis that the model (1) has cointegration rank $r = p$ and k lagged differences. Therefore, the joint identification of r and k, if both are unknown, becomes tricky.

It is also shown that there exists a number of equivalent FCVAR models, which causes problems with identification of fractional parameters d, b and lag length when the cointegration rank r is known, see Carlini and Santucci di Magistris (2017). They demonstrate that for any $k \geq k_0$, where k_0 denotes the number of lagged differences in the true DGP, the following holds:

- Given k_0 and k, with $k \geq k_0$, the number of equivalent sub-models that can be obtained is $m = [\frac{k+1}{k_0+1}]$, where $[x]$ denotes the greatest integer less than or equal to x.
- For any $k \geq k_0$, all the equivalent sub-models are found for parameter values $d_j = d_0 - \frac{j}{j+1} b_0$ and $b_j = b_0/(j+1)$ for $j = 0, 1, ..., m-1$.
- α, β are the same in these models.

Further, they give the number of equivalent sub-models in the following table

$k_0 \downarrow k \rightarrow$	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-	1	1	2	2	3	3	4	4	5	5	6	6
2	-	-	1	1	1	2	2	2	3	3	3	4	4
3	-	-	-	1	1	1	1	2	2	2	2	3	3
4	-	-	-	-	1	1	1	1	1	2	2	2	2
5	-	-	-	-	-	1	1	1	1	1	1	2	2

Thus there are identification problems regarding the parameters $d, b, \Gamma_j, j = 1, \ldots, k$.

2.2 Avarucci's FECM model

The model of Avarucci (2007), which we denote in this paper as FECM, is given by the following dynamics

$$\Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} B_j L^j \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} A_j L^j \Delta^d X_t + \varepsilon_t \qquad \varepsilon_t \sim iid(0, \Omega) \qquad (3)$$

where $B_j = -A_j(\alpha\beta')$. Avarucci (2007) imposes the restriction $X_t = 0$ for $t < 1$. This model is similar to the model of Lobato and Velasco (2006) for testing for fractional unit root in the univariate framework.

The model of Avarucci (2007) has been derived using a standard assumption in a parametric framework (see Robinson and Hualde (2003), Dueker and Startz (1998) that the dynamics of the stationary process can be given by an autoregressive representation. Consider a fractionally

cointegrated system in a triangular form, i.e.

$$\begin{aligned} \xi' \Delta^d X_t &= u_{1t}, \\ \beta' \Delta^{d-b} X_t &= u_{2t}, \end{aligned} \quad \text{with } d-b \geq 0. \tag{4}$$

The triangular representation (4) can be shown to be equivalent to the FVECM without short run dynamics, i.e.

$$\Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + \xi_t, \tag{5}$$

where $\alpha = -\xi_\perp (\beta' \xi_\perp)^{-1}$ and $\beta' \alpha = -I_r$, and r is the cointegration rank. The process u_t has the $VAR(k)$ representation $A(L)u_t = v_t$. Then ξ_t is also a $VAR(k)$ process, i.e.: $\xi_t = \sum_{j=1}^{k} A_j \xi_{t-j} + \varepsilon_t$. Consider the model (5), then

$$\xi_t = \Delta^d X_t - \alpha \beta' \Delta^{d-b} L_b X_t, \tag{6}$$

can be written as $\Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} A_j \xi_{t-j} + \varepsilon_t$ and further using (6) $\Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} A_j [\Delta^d X_{t-j} - \alpha \beta' \Delta^{d-b} L_b X_{t-j}] + \varepsilon_t$ to give finally $\Delta^d X_t = \alpha \beta' \Delta^{d-b} L_b X_t + \sum_{j=1}^{k} A_j \Delta^d X_{t-j} + \sum_{j=1}^{k} B_j \Delta^{d-b} L_b X_{t-j}] + \varepsilon_t$, where $B_j = -A_j(\alpha \beta')$.

The model (3) can also be written in another form. The representation proposed below is coherent with the representation in Johansen (2008). The model (3) can be reformulated as:

$$\Delta^{d-b} \left(I_p - \sum_{j=1}^{k} A_j L^j \right) \left(\Delta^b I_p - \alpha \beta' L_b \right) X_t = \varepsilon_t. \tag{7}$$

This representation emphasizes the nature of the process. In fact, the FECM model is a series created by connecting two systems: a VAR process identified by the lag polynomial $(I - \sum_{j=1}^{k} A_j L^j)$ and a FCVAR process identified by the lag polynomial $\Delta^{d-b}(\Delta^b I_p - \alpha \beta' L_b)$. The following scheme represents the FECM process:

$$\varepsilon_t \rightarrow \boxed{A(L)^{-1}} \rightarrow V_t \rightarrow \boxed{\Pi_{d,b}(L_b)^{-1}} \rightarrow X_t$$

The input of the system is the Gaussian error term ε_t transformed in a VAR process V_t through

the transfer function $A(L)^{-1}$. Finally the VAR process V_t is transformed into a Fractionally Cointegrated process by means of the transfer function $\Pi_{d,b}(L_b)^{-1} := \Delta^{b-d}(\Delta^b I_p - \alpha\beta' L_b)^{-1}$.

In linear system theory, the dynamics of two systems connected in a series can be analysed by checking the zeros and poles of their transfer functions contemporaneously. Hence, the dynamics of the FECM can be found by checking the characteristic roots of the polynomials $A(z)$ and $\Pi_{d,b}(y)$, where $y = 1 - (1-z)^b$. This means that we generate fractional cointegration if $det(\Pi_{d,b}(y)) = 0$ has some of the characteristic fractional roots equal to one and $\beta'\alpha$ is a full rank matrix.

The FECM model is characterized by a different (and more complicated) lag structure than the model proposed in Granger (1986), i.e.:

$$\Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_{t-1} + \sum_{j=1}^{k} \Gamma_j L^j \Delta^d X_t + \varepsilon_t, \tag{8}$$

and the FCVAR model (1) discussed in the previous section. In fact, FECM model (3) contains both the usual lags based on a standard lag operator present in Granger's model (8) and lags using the fractional lag operator. The latter are different than those present in the FCVAR model (1) of Johansen. However, in the very particular case of $d = b = 1$ with $k = 0$ all three models reduce to the standard ECM. Besides, when $k = 0$, i.e. the short run dynamics components are not present, then FCVAR model (3) and FECM model (1) are equal apart for the initial values. The solution of Johansen's FCVAR model depends on the initial values, for $t < 0$, while the FECM model of Avarucci implicitly has the restriction for which the process starts in $t = 0$.

The moving average representation (MA) of the FECM model (3) is given in Avarucci (2007). Following his Theorem 2.2, X_t has the representation

$$X_t = C\Delta_+^{-d} V_t + C^* \Delta_+^{-d+b} V_t + \Delta_+^{-d+2b} \sum_{j=1}^{t-1} \Phi_j V_{t-j}, \tag{9}$$

where $\sum_{j=0}^{\infty} ||\Phi_j||^2 < \infty$, and $C = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp$, and $C^* = -[\bar{\beta}\bar{\alpha}' + C\bar{\beta}\bar{\alpha}' + \bar{\beta}\bar{\alpha}'C + C\bar{\beta}\bar{\alpha}'C]$, where $\Phi_j, j = 1, \ldots, t-1$ are $p \times p$ matrices, $V_t = A(L)^{-1}\varepsilon_t$ and if c is a generic $p \times r$ matrix then $\bar{c} := c(c'c)^{-1}$ and $c_\perp$ is a $p \times (p-r)$ matrix such that $c'_\perp c = c' c_\perp = 0$. Thus, X_t and $\beta' X_t$ are Type

II $I(d)$ and $I(d-b)$ processes respectively.

The proof of Theorem 2.2 is largely based on Theorem 8 in Johansen (2008) and the MA representation (9) is based on the solution (2) given in Johansen (2008, 2009).

Therefore both models generate the same class of processes. However, in the FECM model proposed by Avarucci (2007), cointegration always occurs if $b > 0$ unlike in the FCVAR model of Johansen (2008, 2009), where the system can not be cointegrated for $b > 0$ if $\alpha\beta'$ is a full rank matrix. Thus, the model of Avarucci (2007) has more natural interpretation of the parameter b.

3 Statistical identification of the Avarucci's FECM model

3.1 The nesting structure of the Avarucci's FECM model

The nesting structure of the Avarucci's FECM model differs from the presented in Section 2 the nesting structure of the Johansen's (2008, 2009) FCVAR model and it follows the simple VAR structure, which makes testing the cointegration rank r and the number of lagged differences k to be straightforward. If we define the model

$$\mathcal{H}_{r,k}: \quad \Delta^{d-b}(I_p - \sum_{i=1}^{k} A_i L^i)(\Delta^b I_p - \alpha\beta' L_b)X_t = \varepsilon_t, \quad r = 0, \ldots, p$$

then, the nesting structure of the Avarucci's FECM model is given by

$$\begin{array}{ccccccccc}
\mathcal{H}_{0,0} & \subset & \mathcal{H}_{0,1} & \subset & \mathcal{H}_{0,2} & \subset & \ldots & \subset & \mathcal{H}_{0,k} \\
\cap & & \cap & & \cap & & & & \cap \\
\mathcal{H}_{1,0} & \subset & \mathcal{H}_{1,1} & \subset & \mathcal{H}_{1,2} & \subset & \ldots & \subset & \mathcal{H}_{1,k} \\
\cap & & \cap & & \cap & & & & \cap \\
\vdots & & \vdots & & \vdots & & \ddots & & \vdots \\
\cap & & \cap & & \cap & & & & \cap \\
\mathcal{H}_{p,0} & \subset & \mathcal{H}_{p,1} & \subset & \mathcal{H}_{p,2} & \subset & \ldots & \subset & \mathcal{H}_{p,k}
\end{array},$$

For example, the inclusion $\mathcal{H}_{2,1} \subset \mathcal{H}_{2,2}$ can be tested by $A_2 = 0$ and the inclusion $\mathcal{H}_{1,1} \subset \mathcal{H}_{2,1}$ can be tested on a rank restriction on the matrix $\alpha\beta'$. Moreover, it is simple to prove that the model $\mathcal{H}_{0,1}$ is not nested in $\mathcal{H}_{2,0}$ because the term $\alpha\beta' L_b$ is zero in $\mathcal{H}_{0,1}$.

The nesting structure of the Johansen FCVAR model is more complicated as discussed in Section 2, see also Carlini and Santucci de Magistris (2017) for the details. The convenient nesting structure of Avarucci (2007) FECM model not only makes the testing procedures straightforward, but also assures identification of the model, contrary to Johansen (2008, 2009), which we discuss in the following sub-sections.

3.2 Identification of FECM when the lag length is unknown

Recall from Section 2 that in the Johansen's (2008, 2009) FCVAR model (1) there exists a number of equivalent models with overspecified lag length. In order to illustrate that this does not happen in the Avarucci's (2007) FECM model let us consider the model with just 2 lags:

$$\mathcal{H}_2 : \Delta^{d-b} \left(I_p - \sum_{j=1}^{2} A_j L^j \right) \left(\Delta^b I_p - \alpha\beta' L_b \right) X_t = \varepsilon_t. \tag{10}$$

where $\mathcal{H}_2$ indicates the model with $k = 2$ in (7).

Let us demonstrate under which restrictions the two sub-models of $\mathcal{H}_2$: the model with 2 lags $\mathcal{H}_2^{(0)}$ and the model with 1 lag $\mathcal{H}_2^{(1)}$

$$\mathcal{H}_2^{(0)} \quad : \quad \Delta^{d_0-b_0} \left(I_p - (I_p + \tilde{A}_1)L + \tilde{A}_1 L^2 \right) \left(\Delta^{b_0} I_p - \tilde{\alpha}\tilde{\beta}' L_{b_0} \right) X_t = \varepsilon_t. \tag{11}$$

$$\mathcal{H}_2^{(1)} \quad : \quad \Delta^{d_1-b_1} \left(I_p - \bar{A}_1 L \right) \left(\Delta^{b_1} I_p - \bar{\alpha}\bar{\beta}' L_{b_1} \right) X_t = \varepsilon_t. \tag{12}$$

can be reparameterized as in Carlini and Santucci de Magistris (2017), which would indicate identification problems.

First note, that in case of a unit root, the sub-model $\mathcal{H}_2^{(0)}$ in equation (11) can be written as:

$$\Delta^{d_0-b_0} \left(I_p - \tilde{A}_1 L \right) \left(\Delta^{b_0} I_p - \tilde{\alpha}\tilde{\beta}' L_{b_0} \right) (I_p - I_p L) X_t = \varepsilon_t.$$

or equivalently as

$$\Delta^{d_0-b_0}\left(I_p-\tilde{A}_1L\right)\left(\Delta^{b_0}I_p-\tilde{\alpha}\tilde{\beta}'L_{b_0}\right)\Delta X_t=\varepsilon_t$$

Therefore,

$$\mathcal{H}_2^{(0)}:\Delta^{d_0-b_0+1}\left(I_p-\tilde{A}_1L\right)\left(\Delta^{b_0}I_p-\tilde{\alpha}\tilde{\beta}'L_{b_0}\right)X_t=\varepsilon_t \tag{13}$$

Now, let us compare the sub-models (12) and (13). It is clear that the equations (13) and (12) reparametrize when $\tilde{\alpha}\tilde{\beta}'=\bar{\alpha}\bar{\beta}'$, $\tilde{A}_1=\bar{A}_1$, $b_1=b_0$ and $d_0-b_0+1=d_1-b_1$. Hence, $\mathcal{H}_2^{(0)}=\mathcal{H}_2^{(1)}$ if and only if $d_0+1=d_1$ and there is a unit root.

Furthermore, note that under the unit root the model $\mathcal{H}_2^{(0)}$ is a sub-model of the model $\mathcal{H}_2$, when we impose the restriction $A_2+A_1-I_p=0$. Instead, the sub-model $\mathcal{H}_2^{(1)}$ is the sub-model of the model $\mathcal{H}_2$ when we impose the restriction $A_2=0$.

Therefore, the parameter b is always identified. In order to rule out the identification problem for the parameter d and autoregressive parameters A_j, we only need to assume that the characteristic polynomial

$$\Pi(z)=\left(I_p-\sum_{j=1}^{k}A_jz^j\right)$$

has roots outside the unit circle, which is already assumed in Avarucci (2007). Therefore the identification problem for the parameter d and autoregressive parameters A_j is not present in the Avarucci's FECM model.

3.3 Lack of identification when $\alpha\beta'=0$

However, the problem of identification can arise when $\alpha\beta'=0$. In this situation, (7) is given by

$$\Delta^d\left(I_p-\sum_{j=1}^{k}A_jL^j\right)X_t=\varepsilon_t$$

and the parameter b is not identified. This particular feature of the model has been used in Lasak (2010) to propose a sup-test for no cointegration and is common for all fractionally cointegrated Vector Error Correction models. This identification issue is also relevant in the FCVAR model

when the number of lags is $k = 0$.

3.4 The proof of identification of the Avarucci's FECM model

In this sub-section we give the proof of identification of the FECM model (3) when the lag length k of the VAR is known. The proof of identification follows the same steps as in Johansen and Nielsen (2012).

Theorem 1. Suppose $b \neq 1$ and $\alpha\beta' \neq 0$, then the parameters $\theta = vec(d, b,\ A_1, \ldots,\ A_k, \alpha, \beta, \Omega)$ in the model

$$\Delta^{d-b}\left(I_p - \sum_{j=1}^{k} A_j L^j\right)\left(\Delta^b I_p - \alpha\beta' L_b\right) X_t = \varepsilon_t$$

are identifiable.

Proof

1. A parametric model is identified when $f_{\lambda_0}(x_t|I_{t-1}) = f_{\tilde{\lambda}}(x_t|I_{t-1})$ implies $\lambda_0 = \tilde{\lambda}$, where $f(x_t|I_{t-1})$ is the conditional density function. In the FECM model the parameter vector is given by $\lambda = vec(d, b, \alpha, \beta, A_1, \ldots, A_k, \Omega)$.

2. In the model (3) ε_t is assumed to be iid. We are interested in the first and the second moment of f. Hence we have to show that the conditions $E_{\lambda_0}(x_t|I_{t-1}) = E_{\tilde{\lambda}}(x_t|I_{t-1})$ and $\text{var}_{\lambda_0}(x_t|I_{t-1}) = \text{var}_{\tilde{\lambda}}(x_t|I_{t-1})$ imply $\lambda_0 = \tilde{\lambda}$. The equality for conditional variances requires that $\Omega_0 = \tilde{\Omega}$.

3. We use the decomposition $I_p = \beta\bar{\beta}' + \beta_\perp\bar{\beta}'_\perp$ where $\bar{\beta}' = (\beta'\beta)^{-1}\beta'$ and $\bar{\beta}'_\perp = (\beta'_\perp\beta_\perp)^{-1}\beta'_\perp$ to identify the parameters α and β defining $\tilde{\tilde{\alpha}} = \alpha\beta'\bar{\beta}_0$ and $\tilde{\tilde{\beta}} = \beta(\bar{\beta}'_0\beta)^{-1}$ so that $\alpha\beta' = \tilde{\tilde{\alpha}}\tilde{\tilde{\beta}}'$.

4. The equality for conditional means requires that

$$\Pi_{\lambda_0}(z) = (I_p - A_1^0 z - \cdots - A_k^0 z^k)(1-z)^{d_0-b_0}((1-z)^{b_0} I_p - \alpha_0\beta_0'(1-(1-z)^{b_0}) =$$

$$= (I_p - \tilde{A}_1 z - \cdots - \tilde{A}_k z^k)(1-z)^{\tilde{d}-\tilde{b}}((1-z)^{\tilde{b}} I_p - \tilde{\alpha}\tilde{\beta}'(1-(1-z)^{\tilde{b}}) = \Pi_{\tilde{\lambda}}(z)$$

If $k > 0$ and $r > 0$ then it is implied that $A_j^0 = \tilde{A}_j, j = 1, \ldots, k, d_0 = \tilde{d}, b_0 = \tilde{b}, \alpha_0 = \tilde{\alpha}$ and $\beta_0 = \tilde{\beta}$ when $b_0 = \tilde{b} \neq 1$.

5. If $b_0 = \tilde{b} = 1$ then $(I_p - A_1^0 z - \cdots - A_k^0 z^k)((1-z)I_p - \alpha_0 \beta_0'(1-(1-z)) = (I_p - \tilde{A}_1 z - \cdots - \tilde{A}_k z^k)((1-z)I_p - \tilde{\alpha}\tilde{\beta}'(1-(1-z))$ is not generally solved by $A_j^0 = \tilde{A}_j, j = 1, \ldots, k, \alpha_0 = \tilde{\alpha}$ and $\beta_0 = \tilde{\beta}$. In fact, by the theory of matrix polynomials in Dennis at al. (1976), this problem can be easily explained. Suppose that you have a given matrix polynomial

$$B(z) = I_p - B_1 z - \ldots - B_k z^k$$

and $B_j, j = 1, \ldots, k$ are $p \times p$ fixed square matrices and we want to decompose it as

$$(I_p - D_1 z - \ldots D_{k-1} z^{k-1})(I_p - C_1 z)$$

where C_1 is called the right solvent of the matrix polynomial $B(z)$. We define the latent values as the values $z_1, \ldots, z_{pk}$ such that $|B(z_k)| = 0$ and the right latent vectors as the vectors $v_1, \ldots, v_{pk}$ such that $B(z_j)v_j = 0$ where z_j is a latent value. If $B(z)$ has p linearly independent right latent vectors $v_{1,}, \ldots, v_p$ corresponding to latent roots $z_1, \ldots, z_p$, then $C_1 := Q\Lambda Q^{-1}$ is a right solvent, where $Q = [v_1, \ldots, v_p]$ and $\Lambda = diag(z_1, \ldots, z_p)$, see Dennis et al. (1976). Therefore, in general the right solvent is not unique (because we can find many $z_1, \ldots, z_p$ that satisfy the requirement of the theorem) and there exist different matrices $D_j^{(l)}, j = 1, \ldots, k-1, C_1^{(l)}$ for $l = 1, \ldots, p$ that satisfy the decomposition of $B(z)$. For this reason, when $b_0 = \tilde{b} = 1$ the matrices $A_j, j = 1, \ldots, k$ (the $D_j, j = 1, \ldots, k-1$ matrices in the example) and α and β (the C_1 matrix in the example) are not identified.

6. Suppose now $r = 0$, then the parameters $(d, A_1, \ldots, A_k)$ are just identified and it follows the same argument as in Johansen and Nielsen (2012).

This proof shows that an identification problem occurs when the DGP value of the cointegration gap parameter b_0 is equal to 1. This identification issue can naturally affect the asymptotic distributions of the parameters of the model. We will discuss in Section 4.2 the consequences of the identification

issue on the estimation method proposed.

4 Estimation of the Avarucci's FECM model

In this section we use a profile likelihood approach to estimate the parameters of the FECM model. We concentrate the likelihood function on the parameters $\psi = (d, b)'$ as in Johansen and Nielsen (2012). The profile maximum likelihood estimator is

$$\arg\max_{\psi \in \mathcal{K}} \ell_T(\psi), \tag{14}$$

where $\mathcal{K}$ is a compact set defined as $\mathcal{K} = \left\{\eta \leq b \leq d \leq \bar{d}\right\}$ for some values $\eta > 0$ and $\bar{d} > 0$ and

$$\ell_T(\psi) = -\frac{1}{2T}\log\det(\Omega) + \frac{1}{pT}\log(2\pi).$$

Hence, the parameters α, β and $A_j, j = 1, \ldots, k$ are considered as nuisance parameters.

To maximize the likelihood function we use the same idea as in Johansen and Nielsen (2012). For any combination of ψ, we maximize the likelihood function with respect to the nuisance parameters α, β and $A_j, j = 1, \ldots, k$ with a numerical routine based on a switching algorithm described below.

4.1 The switching algorithm

For any given values of $\psi = (d, b)'$, we estimate

$$\Delta^d X_t - \alpha\beta'\Delta^{d-b}L_b X_t + A(I_k \otimes \alpha)(I_k \otimes \beta')\Delta^{d-b}L_b Z_t - A\Delta^d Z_t = \varepsilon_t.$$

Now we can use the switching algorithm to maximize the likelihood with respect to ψ.

Note that

$$\underset{p\times pk}{A} = \begin{bmatrix} A_1 & A_2 & \cdots & A_k \end{bmatrix} \quad \underset{pk\times 1}{Z_t} = \begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-k} \end{bmatrix}.$$

We run the following switching algorithm:

Step 1. For given values of α, β, Ω, we estimate A in the following equation

$$\Delta^d X_t - \alpha\beta'\Delta^{d-b} L_b X_t = A[\Delta^d Z_t - (I_k \otimes \alpha)(I_k \otimes \beta')\Delta^{d-b} L_b Z_t] + \varepsilon_t$$

The parameters $\hat{A}$ are estimated with ordinary least squares.

Step 2. Given values of A, β, Ω, we estimate α in the following equation:

$$\Delta^d X_t - A\Delta^d Z_t = [I_p : -A](I_{k+1} \otimes \alpha)(I_{k+1} \otimes \beta')\Delta^{d-b} L_b \begin{bmatrix} X_t \\ Z_t \end{bmatrix} + \varepsilon_t$$

Using the Mosconi and Paruolo algorithm explained in the Appendix, we estimate α with generalized least squares by imposing $H_\alpha = [I_p : -A]$, $W_t = (I_{k+1} \otimes \beta')\Delta^{d-b} L_b \begin{bmatrix} X_t \\ Z_t \end{bmatrix}$ and K is the matrix such that $(I_{k+1} \otimes \alpha) = Kvec(\alpha)$.

Step 3. For given values of A, α, Ω, we estimate β in the following equation

$$\Delta^d X_t - A\Delta^d Z_t = [I_p : -A](I_{k+1} \otimes \alpha)(I_{k+1} \otimes \beta')\Delta^{d-b} L_b \begin{bmatrix} X_t \\ Z_t \end{bmatrix} + \varepsilon_t$$

Again, the Mosconi and Paruolo algorithm is needed to estimate β with generalized least squares after imposing $H_\beta = [I_p : -A](I_{k+1} \otimes \alpha)$, $W_t = \Delta^{d-b} L_b \begin{bmatrix} X_t \\ Z_t \end{bmatrix}$ where K is the matrix such that $(I_{k+1} \otimes \beta') = Kvec(\beta)$.

Step 4. For given values of A, α, β we estimate Ω as

$$\Omega = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t'$$

and then we evaluate the likelihood.

We iterate Step 1 - Step 2 - Step 3 and Step 4 until convergence. Finally, we optimize the likelihood function (14) with respect to $\psi = (d, b)'$ to calculate the ML estimator $\hat{\psi}$.

4.2 Estimation and Identification issues

As explained in Section 3, the identification issue in the FECM model arises when the DGP value $b_0 = 1$. In particular, the identification issue is relevant for the matrices α, β and $A_j, j = 1, \ldots, k$, because they are shown not to be unique when $b_0 = 1$. We maximize the profile likelihood function with respect to $\psi = (d, b)'$. Hence, the maximum likelihood estimator $\hat{\psi}$ is always identified, but the estimated nuisance parameters $\hat{\alpha}$, $\hat{\beta}$ and $\hat{A}_j, j = 1, \ldots, k$ are not identified.

By simple algebra, we note that if $b_0 = 1$ then the FECM model with k lags is a reparameterization of the FCVAR model where $b = 1$, $d \geq 1$ ($d = d_0 + 1$) and k lags. In fact, the characteristic polynomial of the FECM model when $b = 1$ is given by the following expression

$$\Delta^{d-1}(I_p - \sum_{j=1}^{k} A_j z^j)((1-z)I_p - \alpha\beta' z) = (1-z)^{d-1}\sum_{j=0}^{k+1} \Psi_j z^j$$

while the characteristic polynomial of the FCVAR model when $b = 1$ is given by the following expression:

$$(1-z)^d I_p = \alpha\beta'(1-z)^{d-1}z - \sum_{j=1}^{k} \Gamma_j (1-z)^d z^j = (1-z)^{d-1}\sum_{j=0}^{k+1} \Psi_j z^j.$$

Furthermore, the FCVAR model with $b = 1$ is an identified model. Hence, we could test in the FECM framework the hypothesis $\mathcal{H}_0 : b = 1$. The asymptotic distribution of this hypothesis has to be derived because this is a case of hypothesis testing in which the nuisance parameters are not

identified. A reference that describe in more detail this problem is Hansen (1996). If the hypothesis $\mathcal{H}_0 : b = 1$ is not rejected, then we study the FCVAR in which $d \geq 1$.

In the Monte Carlo experiment, we have generated a FECM model with $b = 1$, $b = 0.99$, $b = 1.01$ and $T = 100,000$ observations. When we estimate the model (setting the initial values of α, β and A to their true values), the switching algorithm converge very slowly and the number of iterations is approximately of an order of 10^9. These non-identified and almost-non-identified FECM models make the proposed estimation procedure very difficult to manage in the proximity of $b_0 = 1$.

4.3 Initial values

Using the switching algorithm it is important to have a good initial guess for parameters α, β and the matrix A. Hence we use as initial guess for these parameters the estimates from the equation

$$\Delta^d X_t - \alpha\beta' \Delta^{d-b} L_b X_t + \Xi \Delta^{d-b} L_b Z_t - A\Delta^d Z_t = \varepsilon_t$$

obtained with Conditional Sum of Squares (CSS) profile likelihood method. We get $\hat{d}, \hat{b}, \hat{\alpha}, \hat{\beta}, \hat{A}, \hat{\Xi}$ by maximizing the profile likelihood function (or profile CSS) with a reduced rank regression. In fact, this model can be estimated with a profile likelihood depending on the parameters d, b. Hence, we maximize

$$\ell(\psi) = -\log\det\left(\frac{1}{T}\sum_{i=1}^{T} \hat{\varepsilon}_t(\psi)\hat{\varepsilon}_t'(\psi)\right) - \frac{1}{pT}\log(2\pi)$$

over a compact parameter set $\mathcal{K} = \left\{\eta \leq b \leq d \leq \bar{d}\right\}$ for some values $\bar{d} > \eta > 0$.

The estimates $\hat{A}$, $\hat{\alpha}$ and $\hat{\beta}$ are used as initial guess to start up the switching algorithm. The estimates $\hat{d}$ and $\hat{b}$ are used as initial values in the optimization routine.

5 Simulation experiment

The Monte Carlo exercise is conducted with a simulation of the FECM model using the Jensen (2014)'s algorithm to generate the FCVAR model two times. The FECM model can be generated

in two steps: in the first step invert a FCVAR$_{d,b}$ model with 0 lags to obtain Y_t, given by

$$Y_t = (I_p - A_{1,0}L - \ldots - A_{k,0}L^k)^{-1}\varepsilon_t.$$

In the second step generate the process

$$X_t = (\Delta^{d_0} I_p - \alpha_0\beta_0'\Delta^{d_0-b_0}L_{b_0})^{-1}Y_t$$

again with the Jensen (2014) algorithm. We developed a new routine that transforms the parameters of the VAR model into the parameters of the FCVAR model when $d = b = 1$. With these transformed parameters, we generate the process X_t.

We run three Monte Carlo experiments. In the first experiment we generate $N = 1500$ Monte Carlo replications of the following data generating process

$$(I_p - A_{1,0}L - A_{2,0}L^2)(\Delta^{d_0} I_p - \alpha_0\beta_0'\Delta^{d_0-b_0}L_{b_0})X_t = \varepsilon_t$$

where $\varepsilon_t \sim i.i.d.N(0, \Omega_0)$ and $t = 1, \ldots, T$. The numerical parameters of the DGP are

$$A_0 = [A_{1,0} : A_{2,0}] = \begin{bmatrix} -0.2 & 0.2 & 0.2 & 0 \\ 0 & 0.3 & -0.3 & -0.3 \end{bmatrix}$$

$$\alpha_0 = \begin{bmatrix} -0.3 \\ 0.3 \end{bmatrix} \quad \beta_0 = \begin{bmatrix} 1 \\ -0.4 \end{bmatrix} \quad \Omega_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and the values of d_0 and b_0 are chosen such that the inverse roots y of the determinant $|(1-y)I_p - \alpha_0\beta_0'y| = 0$ are outside the fractional circle $\mathbb{C}_{b_0}$ as described in Johansen (2008). The parameters in A_0 are chosen such that the inverse roots z of the lag polynomial calculated as $|I_p - A_{1,0}z - A_{2,0}z^2| = 0$ are outside the unit circle. Furthermore, we have chosen parameters for which no identification issues occur.

For each run of the Monte Carlo simulation we fit the FECM model with two lags, given by

$$(I_p - A_1 L - A_2 L^2)(\Delta^d I_p - \alpha\beta' \Delta^{d-b} L_b) X_t = \varepsilon_t \tag{15}$$

using the switching algorithm discussed in Section 4.1.

When we maximize the likelihood function, the first experiment is to use as initial values for the parameters d and b the true data generating process values d_0 and b_0, while the initial values in the switching algorithm for α, β and Ω are imposed to be α_0, β_0 and Ω_0. The maximization routine climbs the likelihood function within the values $d \in [0.01, 2]$ and $b \in [0.01, 2]$ in order to avoid negative - or close to zero - $\hat{d}$ and $\hat{b}$ estimates. Further, we impose in the maximization routine the restriction $d \geq b$ because in the FECM this inequality must be satisfied.

We introduce two parameters to control the convergence of the switching algorithm. The first parameter is the maximum number of iterations of the switching algorithm N^{iter}. The program stops when a number of N^{iter} chosen is reached. The second parameter is a tolerance number Tol. The program stops when the absolute value of the likelihood at step $k+1$ minus the likelihood value at step k is less than Tol. In the Monte Carlo we have set these two parameters to be $N^{iter} = 20,000$ and $Tol = 10^{-8}$.

To simplify the exposition, we introduce new notation for the elements of the matrices in model. The elements of the matrices are

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \quad \beta = \begin{bmatrix} 1 \\ \beta_1 \end{bmatrix} \quad \Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} \quad A = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{11}^{(2)} & a_{12}^{(2)} \\ a_{21}^{(1)} & a_{22}^{(1)} & a_{21}^{(2)} & a_{22}^{(2)} \end{bmatrix}$$

where $\omega_{12} = \omega_{21}$. When we estimate the FECM model the vectors $\hat{\beta}^* = [\beta_1^* : \beta_2^*]'$ and $\hat{\alpha}^* = [\alpha_1^* : \alpha_2^*]'$ are normalized by calculating $\hat{\alpha} = \beta_1^* \hat{\alpha}^*$ and $\hat{\beta} = \frac{1}{\beta_1^*} \cdot \hat{\beta}^*$.

We present the Monte Carlo results when we simulate the process with $d_0 = 0.8$, $b_0 = 0.6$ and $T = 100,000$. The results of the sample statistics of the distributions of the estimated parameters are reported in Table 3.1. Figure 3.1 displays the plots of the densities of the estimates of the parameters in Eq. 15. These densities are calculated with a non-parametric method and they are

smoothed by a Gaussian Kernel.

The biases of the Monte Carlo estimates are smaller than an order of magnitude of 10^{-3} and the standard deviations are smaller than 0.1. We analyzed if the Monte Carlo sample distributions were normal with a Jarque-Bera test and we do not reject the hypothesis for some of the elements in the matrix A and the fractional parameters d and b.

We tried to check if the switching algorithm is robust with respect to different initial values and the results are still the same when the sample size is $T = 100,000$. If the sample size is $T = 10,000$ then the initial values for α, β and Ω are crucial if we want to find the global maximum of the likelihood function.

The algorithm behaves differently depending on the choice of d_0 and b_0. In fact, if we simulate a model where $d_0 = b_0 = 0.9$, we need over 20,000 iterations of the switching algorithm to converge, which is related to the identification of the model discussed in Section 3.

In the second Monte Carlo experiment we fix $d = d_0$ and $b = b_0$ and let the switching algorithm find the estimates of all other parameters. In Figure 3.2 and Table 3.2 the results are shown when the DGP takes values $d_0 = b_0 = 0.6$ and α_0, β_0, A_0 as before and $T = 100,000$. We notice that in this set up we could not reject the null hypothesis of normality for all the parameters but β.

In the third Monte Carlo experiment we fit the model

$$(I - A_1 L - A_2 L^2)(\Delta^d I_p - \Pi \Delta^{d-b} L_b) X_t = \varepsilon_t$$

where

$$\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix},$$

i.e. matrix Π has a full column rank and we run the switching algorithm fixing $d = d_0$ and $b = b_0$. In Figure 3.3 and Table 3.3 the Monte Carlo results are shown when the DGP has $d_0 = b_0 = 0.6$ and α_0, β_0, A_0 as before and $T = 100,000$. We do not reject the null hypothesis of normality for all the estimated parameters with a Jarque-Bera test.

6 Conclusions

In this paper we discuss two fractionally cointegrated models: a FCVAR model proposed in Johansen (2008, 2009) and the FECM model proposed in Avarucci (2007) and Avarucci and Velasco (2010). They both generate the same class of processes, but due to different lag structures their properties differ significantly. The FECM model turns out to be characterised by a more convenient nesting structure, that allows a straightforward way for testing the cointegration rank and the number of lagged differences to be included as short run parameters. Further, the identification problems are far less severe in FECM model than in the FCVAR model. On the other hand, the estimation of FECM is more complicated due to the presence of two different parts that model the short run dynamics and the restriction that relates their parameters. Thus, we propose an estimation procedure, which is based on the suggestion of Carlini and Mosconi (2014) that maximizes the likelihood function using a switching algorithm and the GLS procedure of Mosconi and Paruolo (2014). We prove the asymptotic distribution of the estimators of all the parameters and we illustrate by means of Monte Carlo experiment the performance of our procedure in finite samples. We find that close to the DGPs chosen in the Monte Carlo simulations, the estimated parameters $\hat{d}, \hat{b}, \hat{\alpha}, \hat{A}_j, j = 1, 2$ are normally distributed, whilst $\hat{\beta}$ has a fat-tailed distribution, which confirms the asymptotic theory developed. The solution of the model has been previously derived in Avarucci (2007), while testing for the rank has been discussed in Łasak and Velasco (for cointegration strength >0.5) and Avarucci and Velasco (for cointegration strength <0.5). Therefore our paper fills in the gap for a complete inference based on Avarucci (2007) model.

Appendix A

The following algorithm describes how to estimate a bilinear form with a GLS model. Further details can be found in Mosconi and Paruolo (2014).

Consider the following equation

$$Y_t = H\theta' W_t + \varepsilon_t \quad t = 1, \ldots, T$$

$$\varepsilon_t \sim iidN(0,\Omega), \quad vec(\theta) = K\psi$$

where Y_t and W_t are respectively $p_y \times 1$ and $p_w \times 1$ vectors, H and θ are respectively $p_y \times r$ and $p_w \times r$ matrices. Then, we can estimate ψ as

$$\hat{\psi} = (K'(H'\Omega^{-1}H \otimes S_{ww})K)^{-1}K'vec(S_{wy}\Omega^{-1}H)$$

where $S_{ww} = \sum_{t=1}^{T} W_t W_t'$ and $S_{wy} = \sum_{t=1}^{T} W_t Y_t'$.

Appendix B

Figure 1: Distributions of the parameters when the DGP parameters are $d_0 = 0.8$, $b_0 = 0.6$. $N = 1,500$ Monte Carlo replications and $Y_t, t = 1, \ldots, 100000$.

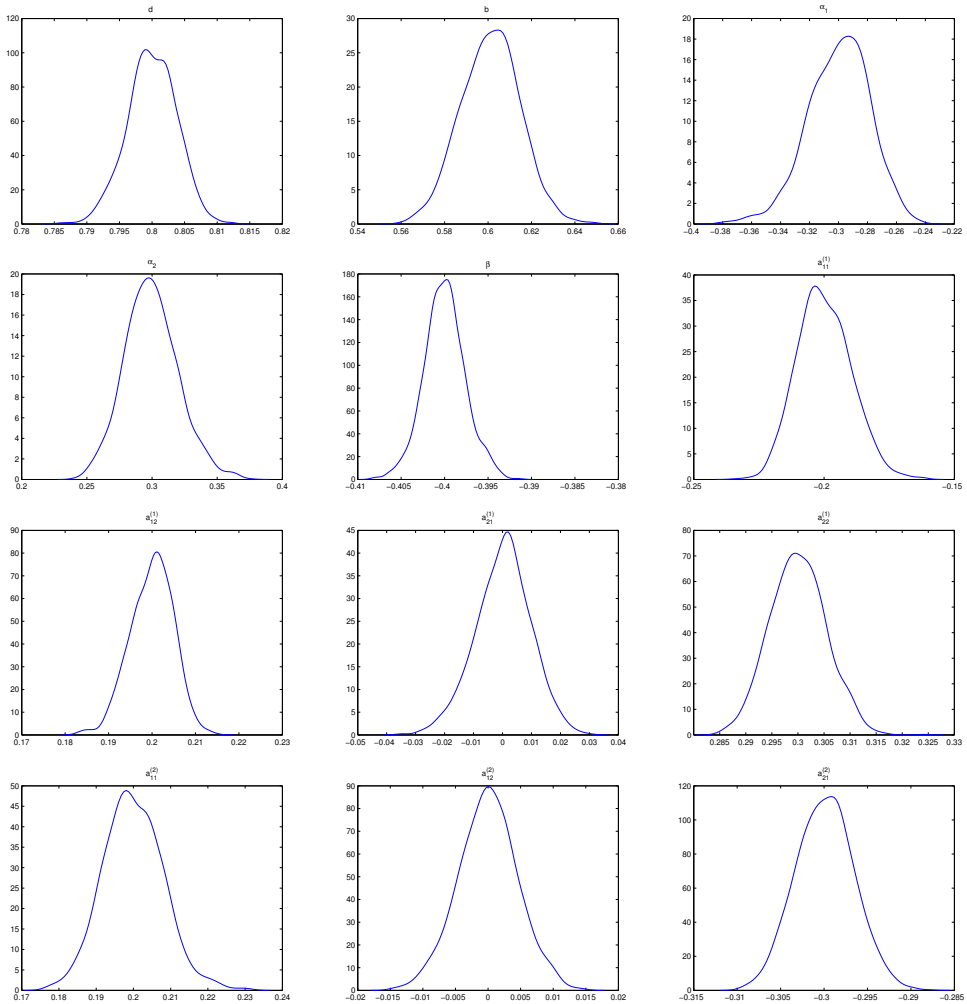


Table 1: Sample statistics of the Monte Carlo distributions when $d_0 = 0.8$ $b_0 = 0.6$, $T = 100000$ observations and $N = 1500$ Monte Carlo replications.

	d	b	α_1	α_2	β
Bias	0.0000	0.0012	0.0001	-0.0005	0.0000
Std.Dev	0.0038	0.0135	0.0214	0.0206	0.0024
Skew.	-0.1404	0.0117	-0.3912	0.3018	0.0546
Kurtosis	3.1239	3.0231	3.3295	3.2352	3.3915
p-value JB test	0.0514	>0.5000	0.0010	0.0010	0.0084
	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$	
Bias	0.0001	0.0000	0.0002	0.0000	
Std.Dev	0.0103	0.0049	0.0096	0.0054	
Skew	0.2826	-0.2777	-0.1961	0.1228	
Kurtosis	3.1502	3.1095	3.1846	3.0136	
p-value JB test	0.0010	0.0010	0.0052	0.1434	
	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$	
Bias	0.0000	0.0000	0.0000	-0.0002	
Std.Dev	0.0081	0.0046	0.0033	0.0030	
Skew	0.2360	-0.0388	0.0532	0.0482	
Kurtosis	3.3815	3.0897	2.9097	2.8717	
p-value JB test	0.0010	>0.5000	>0.5000	0.4295	

Figure 2: Distributions of the Monte Carlo simulations when $d_0 = b_0 = 0.6$ are fixed in the switching algorithm. The generated FECM paths have $T = 100000$ observations.

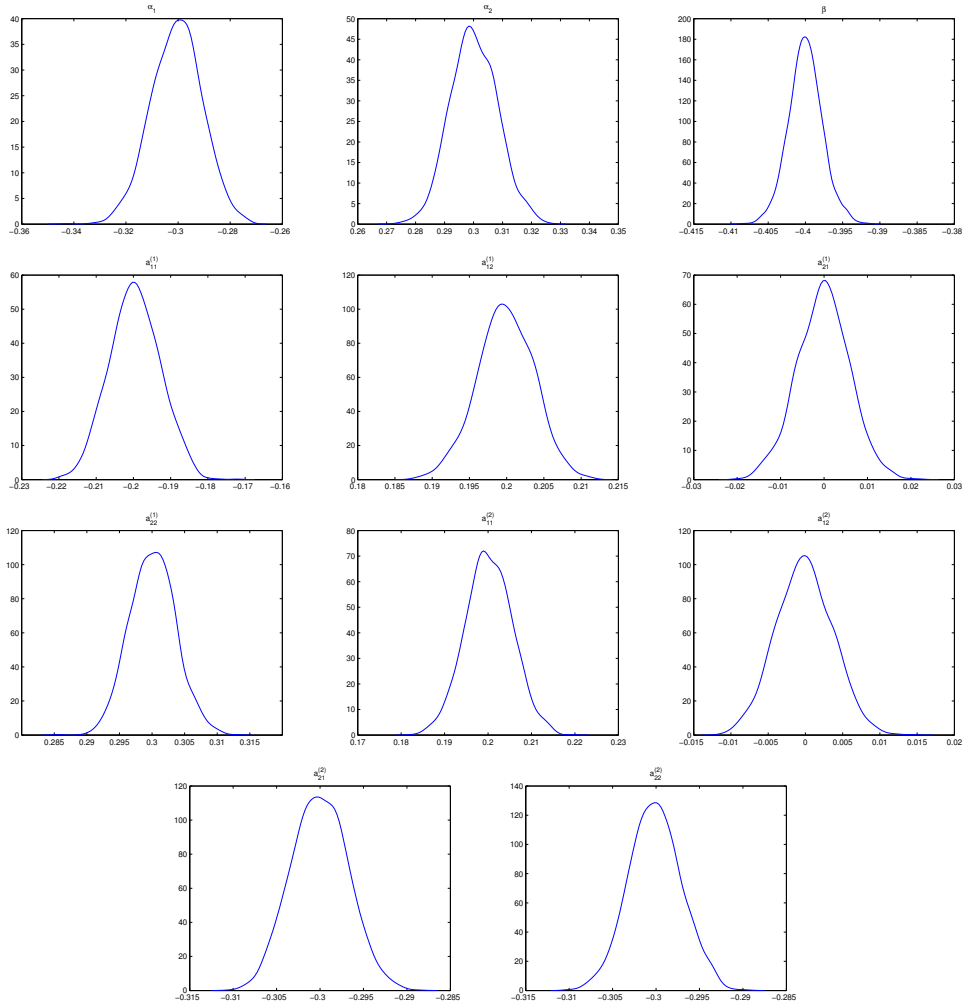


Table 2: Sample statistics of the Monte Carlo distributions $d_0 = b_0 = 0.6$ kept fixed in the switching algorithm with $T = 100000$ and $N = 1750$ Monte Carlo replications.

	d	α_1	α_2	β_1
Bias	-	-0.0005	0.0002	0.0000
Std.Dev	-	0.0097	0.0081	0.0023
Skew.	-	-0.1066	0.0750	0.0655
Kurtosis	-	3.0495	2.9494	3.4033
p-value JB test	-	0.1671	0.3867	0.0031
	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$
Bias	0.0004	-0.0002	-0.0002	0.0001
Std.Dev	0.0068	0.0038	0.0060	0.0035
Skew	0.0432	-0.0834	-0.0087	0.0745
Kurtosis	2.8909	2.9927	3.0914	3.0576
p-value JB test	0.4818	0.3513	>0.5000	0.3810
	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.0002	-0.0001	-0.0000	-0.0001
Std.Dev	0.0054	0.0038	0.0032	0.0030
Skew	0.0200	0.0748	0.0591	0.0374
Kurtosis	2.9412	2.9225	2.8213	2.9423
p-value JB test	>0.5000	0.3455	0.1802	>0.5000

Figure 3: Distributions of the Monte Carlo simulations when $d_0 = b_0 = 0.6$ are fixed in the switching algorithm. The generated FECM paths have $T = 100,000$ observations.

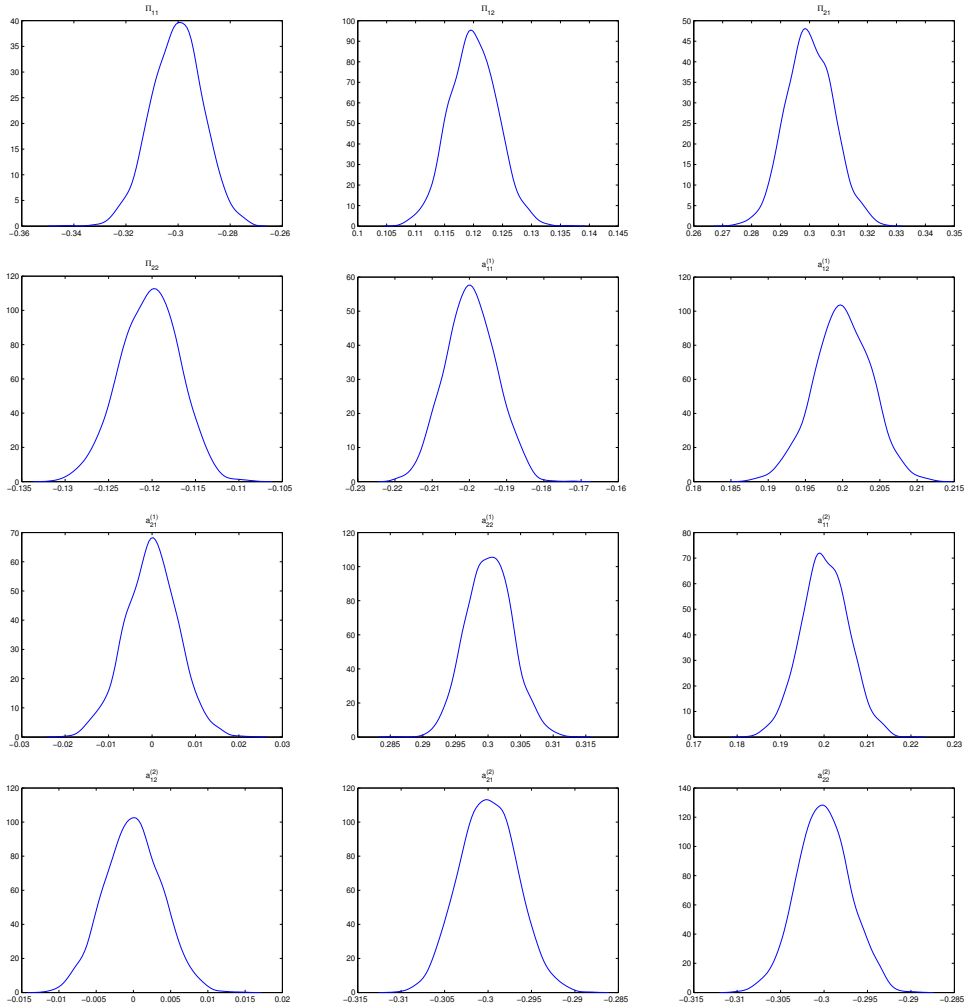


Table 3: Sample statistics of the Monte Carlo distributions $d_0 = b_0 = 0.6$ kept fixed and $T = 100000$, $N = 1750$ Monte Carlo replications and estimation of the matrix Π.

	π_{11}	π_{12}	π_{21}	π_{22}
Bias	- 0.0005	0.0000	0.0002	0.0003
Std.Dev	0.0098	0.0041	0.0081	0.0034
Skew.	-0.1072	0.0678	0.0755	-0.0877
Kurtosis	3.0452	3.0158	2.9507	2.9101
p-value JB test	0.1663	0.4993	0.3846	0.2338
	$a_{11}^{(1)}$	$a_{12}^{(1)}$	$a_{21}^{(1)}$	$a_{22}^{(1)}$
Bias	0.0004	0.0000	-0.0001	0.0002
Std.Dev	0.0068	0.0038	0.0060	0.0035
Skew	0.0428	-0.0699	-0.0070	0.0829
Kurtosis	2.8821	3.0308	3.0953	3.0458
p-value JB test	0.4444	0.4578	$>$0.5000	0.3315
	$a_{11}^{(2)}$	$a_{12}^{(2)}$	$a_{21}^{(2)}$	$a_{22}^{(2)}$
Bias	0.0002	0.0000	0.0000	-0.0001
Std.Dev	0.0054	0.0038	0.0033	0.0030
Skew	0.0202	0.0642	0.0589	0.0379
Kurtosis	2.9403	2.9068	2.8339	2.9455
p-value JB test	$>$0.5000	0.3859	0.2123	$>$0.5000

References

[1] Avarucci, M. (2007), Three Essays on Fractional Cointegration, Ph.D. Thesis, University of Rome Tor Vergata.

[2] Avarucci, M., Velasco C. (2009), A Wald Test for the Cointegration Rank in Nonstationary

Fractional Systems, Journal of Econometrics, 151, 178-189.

[3] Carlini, F., Mosconi, R. (2014), Twice Integrated Models, Manuscript.

[4] Carlini, F., Santucci de Magistris, P. (2017), On the identification of fractionally cointegrated VAR models with the F(d) condition, Forthcoming in Journal of Business and Economics Statistics.

[5] Dennis, J-E., Traub, J.F. and R.P. Weber (1976), The Algebraic Theory of Matrix Polynomials, SIAM Journal on Numerical Analysis, Vol. 13, No.6, pp. 831-845

[6] Dueker, M., Startz, R. (1998), Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates. Review of Economics and Statistics, 83.

[7] Granger, C. W. J. (1986), Developments in the Study of Cointegrated Economic Variables, Oxford Bulletin of Economics and Statistics, 48, 213-28.

[8] Hansen, B.E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, Econometrica 64, 413-430.

[9] Jensen, A. N. (2014), Efficient simulation of the Johansen-Nielsen model, University of Copenhagen, working paper.

[10] Johansen, S. (2008), A representation theory for a class of vector autoregressive models for fractional processes, Econometric Theory, 24, 651-676.

[11] Johansen, S. (2009), Representation of cointegrated autoregressive processes with application to fractional cointegration, Econometric Reviews, 28, 121-145.

[12] Johansen, S., Nielsen, M. O. (2012), Likelihood inference for a fractionally cointegrated vector autoregressive model, Econometrica, 80, 2667-2732.

[13] Łasak, K. (2008), Maximum likelihood estimation of fractionally cointegrated systems, CREATES Research Paper 2008-53.

[14] Łasak, K. (2010), Likelihood based testing for no fractional cointegration, Journal of Econometrics, 158, 67-77.

[15] Lobato, I., Velasco, C. (2006) ,Optimal Fractional Dickey-Fuller tests, Econometrics Journal, 9, 492-510.

[16] Mosconi, R., Paruolo, P. (2014), Rank and order conditions for identification in simultaneous system of cointegrating equations with integrated variables of order two, MPRA Paper 53589,

[17] Robinson, P. M., Hualde, J. (2003), Cointegration in Fractional Systems with Unknown Integration Orders, Econometrica, 71, 1727-1766.

Identification Algorithms Based on the Associative Search of Analogs and Association Rules

Natalia Bakhtadze[1,2], Vladimir Lototsky[1] Valery Pyatetsky[3] Alexey Lototsky[1]

[1]V.A. Trapeznikov Institute of Control Sciences,
Moscow, Russia, sung7@yandex.ru
[2]Bauman Moscow State Technical University, Moscow, Russia
[3]National University of Science and Technology "MISIS",
Moscow, Russia

Abstract. The identification algorithms based on the associative search of analogs and wavelet-analysis are presented. The capabilities of identification algorithms based on the intelligent data analysis to predict the approach of process parameters to critical limits are analyzed, as well as their possible transition to chaotic dynamics. The knowledge base is a set of patterns obtained from the data analysis. The effectiveness of the techniques proposed is demonstrated with the example of product quality prediction in oil refining industry. Resource state prediction methods based on the development of a binary model and a machine learning techniques called association rules are presented

Keywords: Process Identification; Knowledge Base; Associative Search Models; Wavelet Analysis

1 Introduction

One of the main conceptual streams of identification theory and its applications development for a long time is to reduce the uncertainty in the description of the object by means of an adjustable model.

In the statistical description of uncertainty, consistent estimates of plant's characteristics can be obtained by analyzing the convergence of the empirical distribution functional with the corresponding "theoretical" values with sample size increase.

The difficulties in implementing this approach, especially for nonlinear and non-stationary objects, and, on the other hand, increased possibilities of plant history analysis resulted in the development of identification methods based on data mining [1].

In [2-4] the intelligent identification methods, in particular, the employment of additional *a priori* information about the object for system teaching, is outlined as a key trend.

Associative search algorithms based on the design of *virtual* predictive models, which use intelligent analysis of historical process data for dynamic tuning of identification models, can be attributed to this type of methods [3]. The predicting model design

was carried out based on intelligent algorithms of nonlinear dynamic system identification. Those are based on inductive learning: associative search of analogues by intelligent analysis of system parameters archives and the knowledge base developement.

The development of intelligent identification algorithms for nonlinear and unsteady-state objects is important for various applications, in particular, for chemistry, oil refining, and power (smart grids) industries, trading, and transportation systems [2-5].

Issues of the stability of a model built by use of the associative search are considered, in the aspect of the spectrum analysis of the *multi-scale wavelet expansion* [6]. Methods based on the wavelet analysis are characterized with a unique possibility to select "frequency-domain windows".

The construction of predictive model by associative search of dynamic plant on each step is based on system knowledge. This approach allows to use any available *a priori* information about the plant. The criterion of input vectors choice from the archive for current virtual model building underlies clustering technique choice.

Chapter 2 deals with some problems of closed-loop control plant identification models. Chapter 3 is devoted to a discussion of predictive models by means of associative search technique. A case study of this technique is presented in Chapter 4. Conditions of the associative model stability for a non-linear time-varying plant are presented in Chapter 5. Sufficient criteria of approach this plant's dynamics to chaotic one are formulated in Chapter 6. Finally, in Chapter 7 the methods of manufacturing situation's prediction for enterprise resources planning and scheduling are presented based on associative rules.

2 Control system identification

Consider a traditional problem of dynamic object identification. For the inputs meeting Gauss-Markov assumptions, the least squares parameter estimates are consistent, unbiased and efficient [5]. However, building closed-loop control plant models (for identification-based control system synthesis) faces considerable challenges. In a closed loop, the system state depends on control values at earlier time instants. Optimal controllers cause linear state feedbacks resulting in a degenerate problem.

To build an informational model of control system's dynamics in a degenerate case, Moore–Penrose method [7-8] can be used for getting pseudo-solutions to a linear system when the least-squares technique is applied.

3 Associative search technique

There is a variety of processes, which cannot be controlled using linear predictive models. Associative search technique offers [1-3] a constructive solution for nonlinear processes. However, such processes may feature irregularities at certain instants. In

engineering systems, such irregularities often demonstrate oscillating nature. The variability of feed properties due to feed source changes in process industries is a typical example. Another example is seasonal and daily load oscillations in power networks that affect directly the optimization of power transmission control modes. The ups and downs of stock market caused by various economic reasons are also well known. Therefore, the design of predictive models by means of ***associative search*** technique for such type of time-varying processes looks relevant.

Algorithms based on knowledge revealed from process history (inductive knowledge, persistently enriched) implement an intelligent approach to constructing identification models. The intelligence is applying knowledge (*Knowledge Based*) revealed from historical data on the basis of their analysis (*Data Mining*).

The process of knowledge processing in the intelligent system is reduced to recovering (associative search of) *knowledge* over its fragment. Meanwhile, the *knowledge* may be interpreted as associative connections between *images*. As an image, we will use "sets of indicators", that is components of input vectors, input variables.

The criterion of closeness between images may be formulated in very different manners. In the most general case, it may be a logic function, the predicate. In a particular case, when sets of indicators are vectors in *n*-dimensional space, the criterion of closeness may be a distance in this space.

The associative search process may be implemented either as a process of recovering the image over partially given indicators (or recovering a knowledge fragment under the conditions of incomplete information; as a rule, just this process is simulated in different models of the associative memory), or as a process of searching another images that are associatively connected with the given one, respecting to other time instants.

In [4] an approach to form the support of decision making on the control is proposed, based on dynamic modeling the associative search procedure. Results of adoption of the associative search algorithms developed by the authors for industrial processes of the chemical and petroleum manufacturing, processes of control in intelligent power networks (smart grids), trading processes, transport logistic processes.

The method of the *associative search* consists in constructing *virtual* predicting models. The method assumes constructing predicting model of a dynamic plant, being new under each *t*, by use of a set of history data ("associations") formed at the stage of learning and adaptively corrected in accordance to certain criteria, rather than approximating real process in the time. Within the present context, linear dynamic model is of the form:

$$y_N = \sum_{i=1}^{m} a_i y_{N-i} + \sum_{j=1}^{r_s} \sum_{s=1}^{S} b_{j,s} x_{N-j,s}\,, \qquad \forall j = \overline{1,N}, \tag{1}$$

where: y_N is the prediction of the plant output at the time instant N, x_N is the vector of input actions, m is the memory depth in the output, r_s is the memory depth in the input, S is the dimension of the input vectors, a_i and $b_{j,s}$ are the tuned coefficient, meanwhile $x_{N-j,s}$ are selected disregarding the order of the chronological decreasing, have been referred as the *associative pulse.*

Let us note that this model is not classical regression one: there are selected certain inputs in accordance to a certain criterion, rather than all chronological "tail".

The algorithm of deriving the virtual model consists in constructing at each time instant an approximating hypersurface of the space of input vectors and single-dimensional outputs. To construct a virtual model, corresponding to a certain time instant, from the archive, there are selected input vectors being in a certain sense close to the current input vector. An example of selecting the vectors is described below. The dimension of this hyper-surface is selected heuristically. Again, by use of classical (non-recursive) least squares (LS) method there is determined the output value (modeled signal) in the next time instant. Meanwhile, each point of the global non-linear surface of the regression is formed in the result of using linear "'local" models, in each new time instant.

Unlikely to classical regression models, for each fixed time instant from the process history, there are selected input vectors being close to the current input vector in the sense of a certain criterion (rather than the chronological sequence as it is done in regression models). Thus, in equation (1) r_s is the number of vectors from the archive (from the time instant 1 to the time instant N), selected in accordance to the associative search criterion. At each time segment $[N-1, N]$ there is selected a certain set of r_s vectors, $1 \le r_s \le N$. The criterion of selecting the input vectors from the archive to derive the virtual model in the given time instant over the current plant state may be as follows.

Let us introduce as a distance (a norm in $\Re^S$) between points of the *S*-dimensional space of inputs the value:

$$d_{N,N-j} = \sum_{s=1}^{S} \left| x_{N,s} - x_{N-j,s} \right|, \forall j = \overline{1,N}, \tag{2}$$

where $x_{N,s}$ are components of the input vector at the current time instant N.

By virtue of a property of the norm («the triangle inequality», we have:

$$d_{N,N-j} \le \sum_{s=1}^{S} \left| x_{N,s} \right| + \sum_{s=1}^{S} \left| x_{N-j,s} \right|, \qquad \forall j = \overline{1,N}, \tag{3}$$

Let for the current input vector x_N:

$$\sum_{s=1}^{S} \left| x_{N,s} \right| = d_N. \tag{4}$$

To derive an approximating hypersurface for the vector x_N let us select from the archive of the input data such vectors x_{N-j}, $j = \overline{1,N}$ that for a set D_N the condition will hold:

$$d_{N,N-j} \le d_N + \sum_{s=1}^{S} \left| x_{N-j,s} \right| \le D_N, \qquad \forall j = \overline{1,N}, \tag{5}$$

where D_N may be selected, for instance, from the condition:

$$D_N \ge 2d_N^{max} = 2 \max_j \sum_{s=1}^{S} \left| x_{N-j,s} \right|. \tag{6}$$

Under the assumptions that the input actions meet the Gauss-Markov conditions, the estimates obtained via the LS method are unbiased and statistically effective.

4 Oil refining process modeling: a case study

An atmospheric crude distillation process comprises of a crude preheat train, the pre-flash column K-1, atmospheric heaters, the main atmospheric column K-2, and three stripping columns K-6, K-7, and K-9 for atmospheric naphtha (further stabilized and split into lighter and heavier streams in downstream rerun facilities), kerosene, and Diesel products respectively. The atmospheric residuum is further streamed to the vacuum distillation section. To obtain a soft sensor model for the 10% distillation point of the kerosene stream, the lab data for this quality were analyzed along with the process data from the K-2 column. The predictive model is formed by means of the associative search method. The process data were analyzed, and the variables measured both at the plant directly and in the laboratory were selected for modeling. Based on the preliminary data analysis, the following linear predictive model was developed:

$$T(t) = \sum_{i=1}^{4} b_i F_i(t-1) + b_5 F_5(t-3) + b_6 F_6(t-5) + \sum_{i=7}^{12} b_i F_i(t-7),$$

where $T(t)$ is the estimate of the kerosene 10% distillation point; $F_i(t-j)$ are various process parameters, such as flows, temperatures, and pressures, measured directly at the plant, $b_1, \ldots, b_{12}$ are model's coefficients.

The forecast was calculated per linear and associative models for 10,525 time steps (1 step = 10 minutes). Figure 1 shows simulation results for the steps $t = \overline{102, \ldots, 301}$.

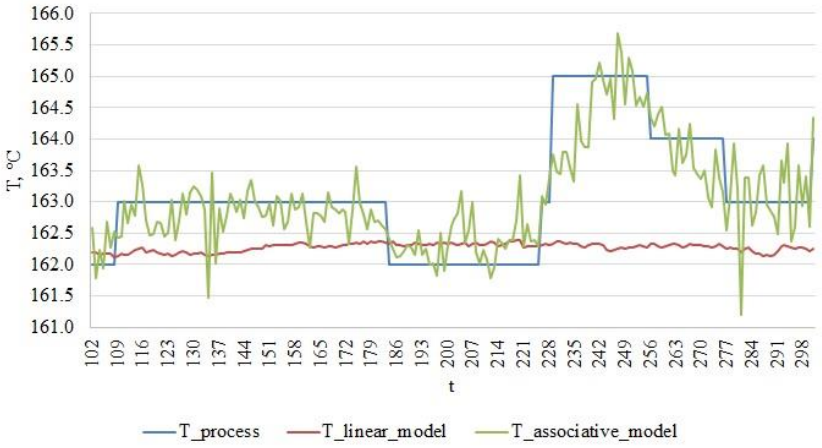


Fig.1. Kerosene 10% distillation point forecast

5 Conditions of the associative model stability in the aspect of the analysis of the spectrum of multi-scale wavelet expansion

Let a predicting associative model of a non-linear time-varying plant meet equation (1). For the selected detail level L for the current input vector $x(t)$, we obtain the multi-scale expansion [6]:

$$x(t) = \sum_{k=1}^{N} c_{L,k}^{x}(t)\varphi_{L,k}(t) + \sum_{l=1}^{L}\sum_{k=1}^{N} d_{l,k}^{x}(t)\psi_{l,k}(t),$$
$$y(t) = \sum_{k=1}^{N} c_{L,k}^{y}(t)\varphi_{L,k}(t) + \sum_{l=1}^{L}\sum_{k=1}^{N} d_{l,k}^{y}(t)\psi_{l,k}7(t), \quad (7)$$

where: L is the depth of the multi-scale expansion ($1 \le L \le L_{max}$, where $L_{max} = \lfloor \log_2 N^* \rfloor$ and N^* is the power of the set of states of the system in the System Dynamics Knowledge Base); $\varphi_{L,k}(t)$ – are scaling functions; $\psi_{l,k}(t)$ are the wavelet functions that are obtained from the mother wavelets by the tension/combustion and shift

$$\psi_{l,k}(t) = 2^{l/2}\psi_{\text{mother}}(2^l t - k)$$

(as the mother wavelets, in the present case we consider the Haar wavelets); l is the level of data detailing; $c_{L,k}$ are the scaling coefficients, $d_{l,k}$ are the detailing coefficients. The coefficients are calculated by use of the Mallat [9].

Let us expand equation (3) over wavelets:

$$\sum_{k=1}^{N} c_{Lk}^{y}(t)\varphi_{Lk}(t) + \sum_{l=1}^{L}\sum_{k=1}^{N} d_{lk}^{y}(t)\psi_{lk}(t) = \sum_{k=1}^{N}\left(\sum_{i=1}^{m} a_i c_{Lk}^{y}(t-i)\varphi_{Lk}(t-i)\right) +$$
$$+\sum_{l=1}^{L}\sum_{k=1}^{N}\left(\sum_{i=1}^{m} a_i d_{lk}^{y}(t-i)\psi_{lk}(t-i)\right) + \sum_{k=1}^{N}\left(\sum_{s=1}^{S}\sum_{j=1}^{r_s} b_{sj} c_{Lk}^{s}(t-j)\varphi_{Lk}(t-j)\right)$$
$$+\sum_{l=1}^{L}\sum_{k=1}^{N}\left(\sum_{s=1}^{S}\sum_{j=1}^{r_s} b_{sj} d_{lk}^{s}(t-j)\psi_{lk}(t-j)\right).$$

Let us consider individually the detailing and approximating parts (8) and (9) correspondingly:

$$(t)\psi_{lk}(t) = \sum_{i=1}^{m} a_i d_{lk}^{y}(t-i)\psi_{lk}(t-i) + \sum_{s=1}^{S}\sum_{j=1}^{r_s} b_{sj} d_{lk}^{s}(t-j)\psi_{lk}(t-j), \quad (8)$$

$$c_{Lk}^{y}(t)\varphi_{Lk}(t) = \sum_{i=1}^{m} \hat{a}_i c_{Lk}^{y}(t-i)\varphi_{Lk}(t-i) + \sum_{s=1}^{S}\sum_{j=1}^{r_s} \hat{b}_{sj} c_{Lk}^{s}(t-j)\varphi_{Lk}(t-j). \quad (9)$$

In [6] it was shown that a sufficient condition of the stability of plant (1) is as follows. For $\forall k = \overline{1,N}$ meeting the inequalities are to be provided:

1. if $m > R$, $R = \max_{s=\overline{1,S}} r_s$, then the conditions for the detailing coefficients are as follows:

$$\left|\frac{a_1 d_{l,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} d_{l,k}^{x_s}(t-1)}{2d_{l,k}^{y}(t)}\right| < 1,$$

$$\left|-\frac{a_2 d_{l,k}^{y}(t-2) + \sum_{s=1}^{S} b_{s,2} d_{l,k}^{x_s}(t-2)}{a_1 d_{l,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} d_{l,k}^{x_s}(t-1)}\right| < 1, \ldots,$$

$$\left|-\frac{a_{R+1} d_{l,k}^{y}(t-R-1)}{a_R d_{l,k}^{y}(t-R) + \sum_{s=1}^{S} b_{s,R} d_{l,k}^{x_s}(t-R)}\right| < 1, \quad (10)$$

$$\left|-\frac{a_{R+2} d_{l,k}^{y}(t-R-2)}{a_{R+1} d_{l,k}^{y}(t-R-1)}\right| < 1, \ldots,$$

$$\left|-\frac{2a_m d_{l,k}^{y}(t-m)}{a_{m-1} d_{l,k}^{y}(t-m+1)}\right| < 1$$

And for the approximating coefficients the conditions are as follows:

$$\left|\frac{a_1 c_{L,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} c_{L,k}^{x_s}(t-1)}{2c_{L,k}^{y}(t)}\right| < 1, \quad (11)$$

$$\left|-\frac{a_2 c_{L,k}^{y}(t-2)+\sum_{s=1}^{S} b_{s,2} c_{L,k}^{x_s}(t-2)}{a_1 c_{L,k}^{y}(t-1)+\sum_{s=1}^{S} b_{s,1} c_{L,k}^{x_s}(t-1)}\right|<1, \ldots,$$

$$\left|-\frac{a_{R+1} c_{L,k}^{y}(t-R-1)}{a_R c_{L,k}^{y}(t-R)+\sum_{s=1}^{S} b_{s,R} c_{L,k}^{x_s}(t-R)}\right|<1,$$

$$\left|-\frac{a_{R+2} c_{L,k}^{y}(t-R-2)}{a_{R+1} c_{L,k}^{y}(t-R-1)}\right|<1, \ldots,$$

$$\left|-\frac{2 a_m c_{L,k}^{y}(t-m)}{a_{m-1} c_{L,k}^{y}(t-m+1)}\right|<1;$$

2. if $m<R,\ R=\max_{s=\overline{1,S}} r_s$, then the conditions for the detailing coefficients are:

$$\left|\frac{a_1 d_{l,k}^{y}(t-1)+\sum_{s=1}^{S} b_{s,1} d_{l,k}^{x_s}(t-1)}{2 d_{lk}^{y}(t)}\right|<1,$$

$$\left|-\frac{a_2 d_{l,k}^{y}(t-2)+\sum_{s=1}^{S} b_{s,2} d_{l,k}^{x_s}(t-2)}{a_1 d_{l,k}^{y}(t-1)+\sum_{s=1}^{S} b_{s,1} d_{l,k}^{x_s}(t-1)}\right|<1, \ldots,$$

$$\left|-\frac{\sum_{s=1}^{S} b_{s,m+1} d_{l,k}^{x_s}(t-m-1)}{a_m d_{l,k}^{y}(t-m)+\sum_{s=1}^{S} b_{s,m} d_{l,k}^{x_s}(t-m)}\right|<1,$$

$$\left|-\frac{\sum_{s=1}^{S} b_{s,m+2} d_{l,k}^{x_s}(t-m-2)}{\sum_{s=1}^{S} b_{s,m+1} d_{l,k}^{x_s}(t-m-1)}\right|<1, \ldots,$$

$$\left|-\frac{2\sum_{s=1}^{S} b_{s,R} d_{l,k}^{x_s}(t-R)}{\sum_{s=1}^{S} b_{s,R-1} d_{l,k}^{x_s}(t-R+1)}\right|<1$$

For the approximating coefficients the conditions are as follows:

$$\left|\frac{a_1 c_{L,k}^{y}(t-1)+\sum_{s=1}^{S} b_{s,1} c_{L,k}^{x_s}(t-1)}{2 c_{L,k}^{y}(t)}\right|<1,$$

$$\left|-\frac{a_2 c_{L,k}^{y}(t-2)+\sum_{s=1}^{S} b_{s,2} c_{L,k}^{x_s}(t-2)}{a_1 c_{L,k}^{y}(t-1)+\sum_{s=1}^{S} b_{s,1} c_{L,k}^{x_s}(t-1)}\right|<1, \ldots,$$

$$\left|-\frac{\sum_{s=1}^{S} b_{s,m+1} c_{L,k}^{x_s}(t-m-1)}{a_m c_{L,k}^{y}(t-m)+\sum_{s=1}^{S} b_{s,m} c_{L,k}^{x_s}(t-m)}\right|<1, \tag{12}$$

$$\left|-\frac{\sum_{s=1}^{S} b_{s,m+2} c_{L,k}^{x_s}(t-m-2)}{\sum_{s=1}^{S} b_{s,m+1} c_{L,k}^{x_s}(t-m-1)}\right|<1, \ldots, \left|-\frac{2\sum_{s=1}^{S} b_{s,R} c_{L,k}^{x_s}(t-R)}{\sum_{s=1}^{S} b_{s,R-1} c_{L,k}^{x_s}(t-R+1)}\right|<1;$$

3. if $m = R \neq 1,\ R = \max_{s=\overline{1,S}} r_s$, then the conditions of the stability for the detailing coefficients are as follows:

$$\left|\frac{a_1 d_{l,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} d_{l,k}^{x_s}(t-1)}{2 d_{l,k}^{y}(t)}\right| < 1,$$

$$\left|-\frac{a_2 d_{l,k}^{y}(t-2) + \sum_{s=1}^{S} b_{s,1} d_{l,k}^{x_s}(t-2)}{a_1 d_{l,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} d_{l,k}^{x_s}(t-1)}\right| < 1, \ldots, \tag{13}$$

$$\left|-\frac{2\left[a_m d_{l,k}^{y}(t-m) + \sum_{s=1}^{S} b_{s,m} d_{l,k}^{x_s}(t-m)\right]}{a_{m-1} d_{l,k}^{y}(t-m+1) + \sum_{s=1}^{S} b_{s,m-1} d_{l,k}^{x_s}(t-m+1)}\right| < 1$$

And for the approximating coefficients the conditions are as follows:

$$\left|\frac{a_1 c_{L,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} c_{L,k}^{x_s}(t-1)}{2 c_{L,k}^{y}(t)}\right| < 1,$$

$$\left|-\frac{a_2 c_{L,k}^{y}(t-2) + \sum_{s=1}^{S} b_{s,2} c_{L,k}^{x_s}(t-2)}{a_1 c_{L,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} c_{L,k}^{x_s}(t-1)}\right| < 1, \ldots, \tag{14}$$

$$\left|-\frac{2\left[a_m c_{L,k}^{y}(t-m) + \sum_{s=1}^{S} \hat{b}_{s,m} c_{L,k}^{x_s}(t-m)\right]}{a_{m-1} c_{L,k}^{y}(t-m+1) + \sum_{s=1}^{S} \hat{b}_{s,m-1} c_{L,k}^{x_s}(t-m+1)}\right| < 1;$$

4. if $m = R = 1,\ R = \max_{s=\overline{1,S}} r_s$, then the condition of the stability for the detailing coefficients is:

$$\left|\frac{a_1 d_{l,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} d_{l,k}^{x_s}(t-1)}{d_{l,k}^{y}(t)}\right| < 1 \tag{15}$$

for the approximating coefficients the condition is:

$$\left|\frac{a_1 c_{L,k}^{y}(t-1) + \sum_{s=1}^{S} b_{s,1} c_{L,k}^{x_s}(t-1)}{c_{L,k}^{y}(t)}\right| < 1. \tag{16}$$

6 Predicting transfer to the chaos

The chaotic system dynamics is characterized by considerable dependence on initial conditions, when as close as needed at the initial time instant trajectories during certain time are diverge by a finite distance. The main characteristics of the chaotic behavior is the speed of divergence of the trajectories defined by the senior Lyapunov exponent, whose value characterizes the degree of non-stability, or the degree of the sensitivity to the initial data. For a linear system with a constant matrix, the senior

Lyapunov exponent is $\chi_1 = \max \mathrm{Re}\lambda_i$, where λ_i are the eigenvalues of the system matrix. In other words, $|\chi_1|$ coincides with the conventional degree of the system stability [10].

Thus, conditions (10), (11) are a sufficient condition of the chaotic dynamics prediction, what is a key condition under implementing phase transfers of technological processes under study.

7 Prediction of Manufacturing Situations

Optimal routine enterprise resources planning and scheduling is currently based on detailed mathematical models of production processes [11]. Rescheduling requires model update subject to the current production information.
Present-day industrial sites feature interrelated multi-variable production processes and sophisticated material flow networks; scheduling at such sites poses nonlinear NP-hard optimization problems.
The state of manufacturing resources should be nevertheless assessed and predicted both to improve control agility and to foresee the situations where schedule execution becomes problematic or impossible. Such situations will be further referred to as incidents.
It may make sense to develop intelligent predictive models describing the overall current state of resources employed to execute all production operations of a specific production process.
The term "production resources" will hereinafter mean the following:

- input flows characterized by formal properties dependent on production specificity
- production equipment $d_{ij,}\ i = 1, \dots, N; j = 1, \dots, M$

and other facilities used for performing the *j-th* operation;

- human resources $h_{ij,}\ i = 1, \dots, H; j = 1, \dots, M$ involved in the *j-th* operation;
- other factors: $f_{ij,}\ k = 1, \dots, N; j = 1, \dots, M$

affecting the *j-th* operation such as energy resources and a variety of formal indices and factors related with the production process.
Production resources may be described differently.

1. Some have qualitative characteristics which take on specific values that may be checked against norms at any moment.
2. The state of others such as certain equipment pieces, may be exclusively either "working" or "not working". Remaining life time may be known or not for such resources. The process historian may however keep failure statistics for a specific equipment piece; maintenance downtime statistics may be also available for a specific piece or similar kind of equipment.
3. One more resource type (including human resources) is not subject to maintenance. In case of outage, such resources should be immediately replaced from the backlog. The replacement process is typically fast; therefore, no values other than 1 (OK) and 0 (not OK) should be assigned to such resource.

Assume a model of a specific manufacturing situation as a dynamic schedule fragment comprises the following components:

$$r_{ij}(t) = \{< C_1 >< C_2 >< C_3 >< C_4 >< C_5 \}_{ijt}, \quad (17)$$

where:

$< C_1 > \stackrel{\text{def}}{=} < ijt >$ is a *resource identifier* including the resource number, the operation number and the time stamp (the number of characteristics may be increased).

Other components of the resource state vector at the time moment t may be represented by a binary code**.**

$< C_2 >$ is the code of the numerical value of a state variable; this code is different for each of the above-listed resource types. $< C_3 >$, $< C_4 >$, and $< C_5 >$ will be discussed further.

Consider the resources whose state may be described by some quantitative characteristic, such as inlet flowrate or temperature for chemical processes or an average equipment failure number. For a specific resource, we assume that the characteristic if its state possesses the values on the half-interval [0; 1) (this half-interval was chosen as an example for simplicity, the results can be easily spread to any other one). This half-interval can be represented as the union [0; 0.5)∪[0.5; 1). We will further correspond the symbols {0; 1} to the left and right half-intervals respectively, namely, 0 to the left half-interval, and 1 to the right one.

Each of the two subintervals can be further split in the same way, and, again, the values 0 and 1 can be assigned to the left and the right parts respectively.

In that way, a finite chain of symbols from {0; 1} has a one-to-one correspondence with a half-interval embedded in [0; 1). For a binary partition a chain of *n* symbols corresponds to a half-interval with the length $\frac{1}{2^n}$.

This way, for each value of a numerical characteristic at the current time moment we obtain a code of zeros and ones. The number of positions, as we show further, will determine the accuracy of prediction. For the resources from the categories 2 and 3, the respective codes will have the same value in all positions (either 1 or 0). $< C_3 >$ is the code of the time before the maintenance end. If a resource is available and operated, the respective code consists of ones. $< C_4 >$ is the code of the time before the equipment piece fails with the probability close to 1 (remaining life).

In the scheduling practice, this time is not less than the operating time. However, resource replacement just during the operation may be sometimes more cost-effective. Moreover, the equipment piece may fail unexpectedly. For resource types from categories 1 and 3, $< C_4 >$ has ones in all positions.

$< C_5 >$ is the time before the scheduled end of the operation. In real-life manufacturing situations, time may be wasted (with the need in schedule update) for the reasons neither stipulated in the production model nor caused by equipment failures.

Generally, it is hardly possible to formalize all such causes of schedule disruption. Therefore, their consolidation as the "remaining plan execution time" is a way to allow for these hidden factors in the production state model.

For the developed binary chain, a forecast may be obtained using data mining techniques. It makes sense to apply the methods named ***association rules search*.** [12]. A forecast of a state described by a binary chain with an identifier can be obtained by revealing the most probable combination of two binary sets of values at

a fixed time instant and at the next instant (a one-step forecast). A more distant prediction horizon is also possible.

8 Conclusions

The predicting identification models built and tuned on the basis of the intelligent data analysis, algorithms of the associative search, have been analyzed. The associative search is implemented by use of the inductive technological knowledge.
The algorithms may be successfully applied to identify non-linear time-varying processes. For these purposes, the multiscale wavelet-expansion is used. On the basis of investigating the dynamics of the coefficients of this expansion, one may predict the approach of process parameters to steady limits. Sufficient conditions of the stability and chaos appearance have been obtained. Examples of using the results obtained to investigate processes of oil refining have been presented, and the benefits of associative search models over linear recurrent models are demonstrated.
This work was supported by the Russian Foundation for Basic Research (Project No. 17-07-00235 A)

References

[1] D. Peretzki, A. and Isaksson, A. Carvalho Bittencourt, and Forsman, K.: Data mining of historic data for process identification. Linköping University Electronic Press, Linköping, Sweeden, http://manualzz.com/doc/8482583/modeling-and-diagnosis-of-friction-and-wear-in-industrial (2014).
[2] Bakhtadze, N., Kulba, V., Lototsky, V. and Maximov, E.: Identification-based approach to soft sensors design/ In: Proceedings of IFAC Workshop of Intelligent Manufacturing Systems, pp. 86-92. Alicante, Spain (2007).
[3] Bakhtadze, N., Maximov, E. and Valiakhmetov, R.: Fuzzy soft sensors for chemical and oil refining processes. IFAC Proceedings Volumes 41(2), 4246-4250, (2008).
[4] Bakhtadze, N., Lototsky, V., Vlasov, S., and Sakrutina, E.: Associative Search and Wavelet Analysis Techniques in System Identification. IFAC Proceedings Volumes 45(16), 1227-1232 (2012).
[5] Bakhtadze, Natalya N. and Ekaterina A. Sakrutina: The Intelligent Identification Technique with Associative Search. International Journal of Mathematical Models and Methods in Applied Sciences 9, 418-431 (2015).
[6] Bakhtadze, N, Sakrutina, E. and Pyatetsky, V.: Predicting Oil Product Properties with Intelligent Soft Sensors. IFAC-PapersOnLine 50 (1), 14632–14637 (2017).
[7] Moore, E.: On the reciprocal of the general algebraic matrix. Bulletin of the American Mathematical Society 26, 394-395 (1920).
[8] Penrose, R.: A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society 51(3), pp. 406-413 (1955).
[9] Mallat, S.: A Wavelet Tour of Signal Processing. 2nd ed. Academic, San Diego, CA, (1999).
[10] Fradkov, A. and Evans, R.: Control of chaos: survey – 1997-2000. IFAC Proceedings Volumes, 35(1), 131-142 (2002).
[11] Al-Otabi, G.A. and Stewart, M.D.: Simulation model determines optimal tank farm design. Oil & Gas Journal, 102(7), 50-55 (2004).
[12] Agrawal, R. and Srikant, F.: Discovery of Association Rules. In: Proceedings of the 20th International Conference on VLDB, pp. 487-499. Santiago, Chile (1994).

The impact of the increased domestic energy prices on the Saudi Arabian economy. Insights from KGEMM.

Fakhri J. Hasanov[1,2,3,]✉, Frederic L. Joutz[3] and Jeyhun I. Mikayilov[1]

[1] King Abdullah Petroleum Studies and Research Center, Riyadh 11672, Saudi Arabia,

[2] Department of Socio-Economic Modeling, Institute of Control Systems, Baku, AZ1141, Azerbaijan

[3] Research Program on Forecasting, Department of Economics, George Washington University, Washington, DC 20052, USA

✉ Correspondence: fakhri.hasanov@kapsarc.org

The ITISE 2018 (International conference on Time Series and Forecasting) topics

Time Series Analysis and Forecasting: *Probabilistic approaches to modeling macroeconomic uncertainties; Nonstationarity*
Econometric and Forecasting: *Real macroeconomic monitoring and forecasting; Advanced econometric methods*
Energy forecasting: *Applications in real problem (finance, transportation, networks, meteorology, ehealth, environment, etc.)*

.

Before using formal models, one should make sure that they reasonably capture the dynamic relationships between variables and sectors. In this regard, existing global modeling platforms tend to have an oversimplified representation of the Saudi Arabian economy. Given their coverage and specification these models may not provide policymakers with confidence in their results and detailed insights necessary.

The KAPSARC Global Energy Macroeconometric Model (KGEMM) is constructed and tested specifically for the KSA. It is designed to address the following issues:

(a) Analyze the effects of different policy choices, such as energy price and fiscal policy changes, on the economy.

(b) Assess the effects of the Saudi Vision 2030 initiatives and evaluate its targets.

(c) Link Saudi Arabia's macroeconomic-energy environment with the global economy/energy markets.

KGEMM is a time-series structural energy-macroeconometric model representing the economy using eight inter-related blocks (Real, Energy, Price, Fiscal, Monetary, External, Wage and Labor Market, Population and Age Cohort). The model contains nearly 320 behavioral equations and identities, and about 700 time series variables, over the period *1980 – 2017*. KGEMM has been built following the LSE tradition testing for stationarity, (vector) equilibrium correction mechanisms, forms of exogeneity, and structural change. *Autometrics (2013)* was used in model/equation selection to follow a consistent approach. When making projections, the model attempts to employ recent developments in the theory of forecasting.

Vision 2030 for the Kingdom of Saudi Arabia is a National Transformation Program. Two critical parts of the Vision focus on fiscal balance issues and domestic energy consumption. An important component involves ending energy subsidies. As Saudi Vision 2030, particularly Fiscal Balance Program, outlined, the domestic energy prices increase (or subsidy cut) is one of the key elements for the KSA economic transformation in terms of both obtaining higher revenues for the government budget and making economic agents more efficient in their energy use.

In this study, we simulate KGEMM to project and analyze the effects of alternative energy reform ending subsidies raising domestic energy prices in the KSA from 2018-2025. We focus particularly on their impact on crude oil and products consumption and key macroeconomic indicators such as: GDP, Non-oil GDP, inflation, and net government budget revenues. The simulations show that energy prices increases have significant effects on the macroeconomic-energy environment of KSA. Our findings provide useful information for policymakers considering alternative development paths to meet the objectives of Vision 2030.

Reference: Doornik, J. A., and Hendry, D. F. (2013). Empirical Econometric Modelling using PcGive: Volume I, 7th edn. London: Timberlake Consultants Press.

Yield Curve Modeling with Macro Factors

An Implementation of the Kim Filter in a Two-Economy Markov Regime Switching State-Space Model

András Bebes, Dávid Tran, László Bebesi

Abstract. We present a two-economy yield curve model with macro factors. We assume two distinct economic regimes and use a Markov regime switching state-space model to create a 5-year forecast for the Hungarian and Euro area factors. A rotated dynamic Nelson-Siegel model is used for yield curve modeling, we apply Gibbs sampling for parameter estimation and implement the Kim filter to determine the factor values.

Keywords: Yield Curve Modeling, Dynamic Nelson-Siegel Model, Markov Regime Switching, Gibbs Sampling

1 Introduction

In this paper we construct a Markov regime switching state-space model. We use this model to simultaneously forecast the Hungarian and the Euro area term structures of interest rates as well as some macroeconomic and financial factors, namely the Hungarian and Euro area inflation (CPI) rates, the Hungarian credit default swap (CDS) curve and the EUR/HUF exchange rate. The two yield curves and the macro factors are adequate for modeling the future cash-flows of outstanding government debt securities as well as instruments to be issued in the future by the Hungarian state. Forecasting cash-flows allows the debt manager to optimize the debt portfolio and conduct various scenario analyses.

The scope of this study extends to the model specification, parameter estimation and examination of the forecasted yield curves and macro-financial factors. Pricing the outstanding and future debt instruments as well as portfolio optimization is outside the scope of this paper.

Modeling and forecasting the term structure of interest rates (yield curve) is a well-researched topic. The most popular class of models used in empirical applications are variants of the Nelson-Siegel [1] model. A dynamic extension of the original static model was proposed by Diebold and Rudebusch [2]. We model the yield curve using a rotated dynamic Nelson-Siegel Model following the work of Nyholm [3].

State-space models are useful for modeling dynamic linear systems driven by unobservable parameters. These parameters can be estimated using the Kalman filter [4]. In practice, however, economic applications often exhibit nonlinearity. Hamilton [5] was the first to apply Markov regime switching models to economic problems with the purpose of dealing with time varying parameters. The unobservable parameters of Markov

adfa, p. 1, 2011.

regime switching models can be estimated using an extension of the Kalman filter by Kim [6]. Our work follows Kim and Nelson [7] in using Gibbs sampling for parameter estimation and the Kim filter for determining the factor values.

Using a regime switching model is easily justified by economic considerations. Economic cycles are prevalent and can be separated into longer periods of growth (expansion) and shorter crisis periods of contraction (recession). These two distinct empirical situations allow for a meaningful application of time varying parameters. The steady state values (long-run expectations) determining the factors, the transition matrix and the error terms are all different based on the given regime. The heteroscedasticity of the error terms of the transition equation driving the factors can be handled using a different covariance matrix for every state.

Using a two-economy model with additional macro factors is also necessary as the Hungarian debt portfolio includes foreign currency denominated bonds as well as inflation-linked securities. Although the portfolio includes securities denominated in several currencies, the entire foreign currency exposure is only Euro denominated due to cross-currency swaps.

We can also model the dynamics between a small open economy (Hungary) and a large economy (the Euro area). Our model allows factor dynamics to be specified in a way that the Euro area factors can influence the Hungarian factors but not vice versa.

The structure of this paper is as follows: Section 2 deals with the specification of the used Markov regime switching model. Section 3 gives an overview of the process of parameter estimation. Section 4 introduces the data used for fitting the model. Section 5 gives an analysis of the forecast results. Section 6 concludes.

2 The Markov Regime Switching State-Space Model

In this section we construct a Markov regime switching state-space model that can be used to model the entirety of the Hungarian government debt portfolio and explore the relationship between the examined factors in a coherent and realistic way.

We use a 3-Factor Nelson-Siegel Model to estimate the HUF yield curve, while we assume that the Hungarian Euro yield curve at a specific maturity τ is the sum of the risk-free Euro yield curve and the Hungarian Credit Default Swap curve at maturity τ. We use a 3-Factor Nelson-Siegel Model to estimate the risk-free Euro yield curve.

Following Nyholm [3] we rotate the level and slope factors of the Nelson-Siegel Model to get a more transparent structure for our factors. Thus our level parameter reflects the level of the short-term rate, and the slope factor is positive in case of a normal (upward sloping) yield curve. In accordance with our assumptions the HUF ($y_{t,H}(\tau)$) and the risk-free Euro ($y_{t,E}(\tau)$) yield curve at time t and maturity τ is given by

$$y_{t,H}(\tau) = L_{t,H} + S_{t,H}\left(1 - \frac{1-e^{-\tau/\lambda_{t,H}}}{\tau/\lambda_{t,H}}\right) + C_{t,H}\left(\frac{1-e^{-\tau/\lambda_{t,H}}}{\tau/\lambda_{t,H}} - e^{-\tau/\lambda_{t,H}}\right) + \varepsilon_{\tau,H}, \quad (1)$$

$$y_{t,E}(\tau) = L_{t,E} + S_{t,E}\left(1 - \frac{1-e^{-\tau/\lambda_{t,E}}}{\tau/\lambda_{t,E}}\right) + C_{t,E}\left(\frac{1-e^{-\tau/\lambda_{t,E}}}{\tau/\lambda_{t,E}} - e^{-\tau/\lambda_{t,E}}\right) + \varepsilon_{\tau,E}, \quad (2)$$

where $\varepsilon_{\tau,H} \sim N(0, \quad \sigma^2_{\tau,H})$ and $\varepsilon_{\tau,E} \sim N(0, \quad \sigma^2_{\tau,E})$, $t = 1, \dots, n$. One can easily see that $\lim_{\tau \to 0} y_{t,k}(\tau) = L_{t,k}$ and $\lim_{\tau \to \infty} y_{t,k}(\tau) = L_{t,k} + S_{t,k}$ for $k \in \{H, E\}$. The curvature factor remains unchanged.

To avoid negative CDS prices we model the natural logarithm of the CDS curve with a 2-Factor Nelson-Siegel Model. The Hungarian CDS curve ($y_{t,CDS}(\tau)$) at time t and maturity τ is given by

$$\log\left(y_{t,CDS}(\tau)\right) = L_{t,CDS} + S_{t,CDS}\left(\frac{1-e^{-\tau/\lambda_{t,CDS}}}{\tau/\lambda_{t,CDS}}\right) + \varepsilon_{\tau,CDS}, \quad (3)$$

where $\varepsilon_{\tau,CDS} \sim N\left(0, \quad \sigma^2_{\tau,CDS}\right)$, $t = 1, \dots, n$.

Given our expectations of the system we also incorporate the Hungarian and Euro-zone inflation rates as well as the EUR/HUF exchange rate into our model. The latter is required to enable us to convert forecasted Euro cash-flows into a HUF basis. We observe the macro factors present in our model with measurement error, thus

$$mac_{t,j} = \widehat{mac}_{t,j} + \varepsilon_{mac_j}, \quad (4)$$

where $\varepsilon_{mac_j} \sim N\left(0, \quad \sigma^2_{j,mac}\right)$, $j \in \{cpi_{hun}, cpi_{eu}, eur/huf\}$. Given equations (1)-(4) we have a total of 11 factors.

Let $Y_t = \left[y_{t,H}, y_{t,E}, mac_{t,cpi_{hun}}, mac_{t,\mathrm{cpi}_{eu}}, \Delta\log\left(mac_{t,eur/huf}\right), \log\left(y_{t,CDS}\right)\right]'$ be the vector of our observations[1] while $X_t = \left[L_{t,H}, S_{t,H}, C_{t,H}, L_{t,E}, S_{t,E}, C_{t,E}, mac_t, L_{t,CDS}, S_{t,CDS}\right]$ is the vector of unobserved state variables i.e. the factors at time t ($t = 1, \dots, n$).

Let S_t denote an unobserved, discrete-valued 2-state Markov-switching variable ($S_t \in \{1,2\}$) reflecting our prior assumption that the economy switches between two distinct regimes. The first regime (normal regime) represents economic stability and growth, while the second one (crisis regime) is connected to a more turbulent crisis situation.

The transition probabilities of the 2-state Markov process are given by

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}, \quad (5)$$

where $p_{ij} = Pr(S_t = j | S_t = i)$ and $p_{i1} + p_{i2} = 1$ for all $i \in \{1,2\}$.

The state-space representation of our dynamic linear model at time t can be written as:

$$Y_t = H_t X_t' + \theta_t, \quad (6)$$

[1] Both $y_{t,H}$ and $y_{t,E}$ contain maturities for $\tau \in \{1M, 2M, 3M, 6M, 9M, 12M, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y, 12Y\}$, while $y_{t,CDS}$ contains maturities for $\tau \in \{6M, 12M, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y\}$.

$$X_{t+1} = \mu_{S_t} + X_t F_{S_t} + \varepsilon_t, \tag{7}$$

$$\begin{bmatrix}\theta_t \\ \varepsilon_t'\end{bmatrix} \sim N\left(0, \begin{pmatrix} R_{S_t} & 0 \\ 0 & Q_{S_t}\end{pmatrix}\right) \tag{8}$$

with $t = 1, \dots, n$. The measurement equation (6) describes the relation between the data and our factors, where Y_t, θ_t are $m \times 1$ vectors, X_t, ε_t are $1 \times f_n$ vectors, m determines the dimension of the measurement variables, while f_n accounts for the number of factors in the model[2]. H_t is an $m \times f_n$ matrix that links the observed variables to the unobserved factors at time t, mainly containing the factor-loadings of equations (1)-(3). H_t does not depend on the Markov variable S_t, it is only dependent on the time-decay parameters, $\lambda_{t,k}$ $k \in \{H, E, CDS\}$.

The transition equation (7) describes the dynamics of the state variables at time t, where F_{S_t} $(f_n \times f_n)$ is the transition matrix describing the evolution of our factors, while μ_{S_t} represents the drift component in the transition equation, both are dependent on the Markov-switching variable S_t. R and Q are covariance matrices with appropriate dimensions. The (i, j)-th element of the parameters dependent on S_t can be described by the equation

$$f^k_{i,j,S_t} = f^k_{i,j,1} S_{1t} + f^k_{i,j,2} S_{2t}, \tag{9}$$

where k denotes the chosen parameter ($k \in \{\mu, F, R, Q\}$), with $S_{lt} = 1$ if $S_t = l$, and $S_{lt} = 0$ otherwise for $l \in \{1,2\}$.

Let C be the vector of the steady state factor values, given by $C_{S_t} = \mu_{S_t}\left(I_{f_n} - F_{S_t}\right)^{-1}$, where I_{f_n} is the identity matrix with dimensions $(f_n \times f_n)$. Extracting it from equation (7) and denoting $X^*_t = X_t - C_{S_t}$ we get $X^*_t = X^*_t F_{S_t} + \varepsilon_t$, which we can use to generate samples for C in each step of the Gibbs sampler from the posterior distribution suggested by Villani [8], and thus we can replace μ from our parameter set. Considering that the factors are easy to interpret in econometrics, the steady state values hold valuable information. While an analyst does not necessarily have expectations for a given element of the drift parameter μ, they may have one for the Eurozone CPI.

We have two additional restrictions based on our assumptions. First, R_{S_t} is diagonal, thus at every time t $(t = 1, \dots, n)$ the estimation error in our yield curve points and the measurement errors in the macro factors are uncorrelated. Second, we assume that the foreign factors can influence the domestic ones but not vice versa.

3 Parameter Estimation

This section deals with the problem of estimating the parameters of the system described in Section 2. We call the parameters H, R, C, F and Q the model's hyperparameters, while the entirety of our parameter space is denoted by $\Theta \in$

[2] The index in f_n is in no connection with the count of the observations; n.

$\{H_j, F_i, C_i, Q_i, R_i, \Pr, S_j\}$ $(i \in \{1, 2\}, j \in (1, \ldots, n))$. To estimate the values of the factors we turn to the work of Kim [6], while for the random number generation of the models hyperparameters we use the distribution suggested by Villani [8] and Sugita [9] within the framework of a Gibbs sampler.

3.1 Kim Filter

The Kim filter is an extension of the Kalman filter [4] which gives a solution to the parameter estimation problem of general dynamic linear models with a Markov-switching parameter put in a state-space representation form.

Suppose that the hyperparameters of the model defined in Section 2 are known. Let $\mathcal{F}_t$ be the information available at time t. As defined in the Kalman filter the goal is to forecast X_t based on information available at time $t-1$. Let

$$X_{t|t-1}^{(i,j)} = E(X_t|\mathcal{F}_{t-1}, S_t = j, S_{t-1} = i) \tag{10}$$

be the conditional expectation of X_t based on $\mathcal{F}_{t-1}$ and $S_t = j, S_{t-1} = i$, and

$$P_{t|t-1}^{(i,j)} = E\left(\left(X_t - X_{t|t-1}^{(i,j)}\right)' \left(X_t - X_{t|t-1}^{(i,j)}\right) \middle| \mathcal{F}_{t-1}, S_t = j, S_{t-1} = i\right) \tag{11}$$

the mean squared error of the forecast. Then, conditional on $S_t = j, S_{t-1} = i$ at time t, the following 7 equations define the Kalman filter algorithm; the first 4 define the phase of prediction, the last 2 the phase of updating, while the fifth defines the Kalman gain:

$$X_{t|t-1}^{(i,j)} = \mu_j + X_{t-1|t-1}^{(i)} F_j, \tag{12}$$

$$P_{t|t-1}^{(i,j)} = F_j' P_{t-1|t-1}^{(i)} F_j + Q_j, \tag{13}$$

$$\eta_{t|t-1}^{(i,j)} = Y_t - H_t \left(X_{t|t-1}^{(i,j)}\right)', \tag{14}$$

$$f_{t|t-1}^{(i,j)} = H_t P_{t|t-1}^{(i,j)} H_t' + R_j, \tag{15}$$

$$K_t^{(i,j)} = P_{t|t-1}^{(i,j)} H_t' \left(f_{t|t-1}^{(i,j)}\right)^{-1}, \tag{16}$$

$$X_{t|t}^{(i,j)} = X_{t|t-1}^{(i,j)} + \left(K_t^{(i,j)} \left(\eta_{t|t-1}^{(i,j)}\right)'\right)', \tag{17}$$

$$P_{t|t}^{(i,j)} = \left(I_{f_n} - K_t^{(i,j)} H_t\right) P_{t|t-1}^{(i,j)}. \tag{18}$$

$X_{t|t}^{(i)} = E(X_t|\mathcal{F}_t, S_t = i)$ is the conditional expectation of X_t based on $\mathcal{F}_t$ and $S_t = i$, while $P_{t|t}^{(i)} = E\left(\left(X_t - X_{t|t}^{(i)}\right)' \left(X_t - X_{t|t}^{(i)}\right) \middle| \mathcal{F}_t, S_t = i\right)$ is the mean squared error of X_t conditional on $\mathcal{F}_t$ and $S_t = i$. $\eta_{t|t-1}^{(i,j)}$ is the conditional forecast error of Y_t based on $\mathcal{F}_{t-1}$ and $S_{t-1} = i$, $S_t = j$, while $f_{t|t-1}^{(i,j)}$ is the conditional covariance of $\eta_{t|t-1}^{(i,j)}$.

Assuming that X_t is stationary, the unconditional mean and covariance matrix and the steady state probabilities can be used to start the filter.

To handle the increase in the number of cases at each iteration caused by the different states, Kim suggests the following equations to collapse the terms $X_{t|t}^{(i,j)}$ and $P_{t|t}^{(i,j)}$ to $X_{t|t}^{(j)}$ and $P_{t|t}^{(j)}$ respectively

$$X_{t|t}^{(j)} = \frac{\sum_{i=1}^{M} \Pr(S_t = j, S_{t-1} = i|\mathcal{F}_t) X_{t|t}^{(i,j)}}{\Pr(S_t = j|\mathcal{F}_t)}, \tag{19}$$

$$P_{t|t}^{(j)} = \frac{\sum_{i=1}^{M} \Pr(S_t = j, S_{t-1} = i|\mathcal{F}_t)\left(P_{t|t}^{(i,j)} + \left(X_{t|t}^{(j)} - X_{t|t}^{(i,j)}\right)'\left(X_{t|t}^{(j)} - X_{t|t}^{(i,j)}\right)\right)}{\Pr(S_t = j|\mathcal{F}_t)}. \tag{20}$$

Using these approximations, the problem becomes feasible. To approximate the probability terms in equations (19) and (20) one can use the Hamilton filter suggested by Kim and Nelson [7], thus making the filter complete.

3.2 Estimation of The Model Hyperparameters

With the assumption that H, R, C, F and Q parameters are known for both regimes, the Kim filter is applicable to estimate the values of X_t, and the $P(S_t = j|\mathcal{F}_t)$ probabilities, while S_t can be acquired through random number generation for all $t = 1, \dots, n$. For the estimation of parameters R, C, F and Q for a given X we use Gibbs sampling.

Assuming that $vec(F_{prior})$ is multidimensional normal with $vec(F_0)$ mean and V_0 covariance matrix, while Q_{prior} is inverse-Wishart with Q_0 scale matrix and n_0 degrees of freedom, Sugita [9] has shown that

$$vec(F_{posterior})|Q, Y \sim N(vec(F^*), \quad V), \tag{21}$$

$$Q_{posterior}|F, Y \sim IW(Q^*, \quad n^*), \tag{22}$$

with $V = \left(V_0^{-1} + Q^{-1} \otimes (X'X)\right)^{-1}$ and $vec(F^*) = V\left(V_0^{-1} vec(F_0) + (Q^{-1} \otimes I_m)^{-1} vec(X'Y)\right)$, while $Q^* = (Y - XF)'(Y - XF) + Q_0$, $n^* = n + n_0$, where V and V_0 are covariance matrices with appropriate dimensions $vec(\cdot)$ denotes the vectorization function, $\otimes$ the Kronecker product.

Assuming that C_{prior} is multidimensional normal with c_0 mean and Σ_0 covariance matrix, Villani [8] has shown that the conditional distribution $C_{posterior}$ is also multidimensional normal

$$C_{posterior}|Q, F, X^* \sim N(c^*, \quad \Sigma^*), \tag{23}$$

with $\Sigma^* = (U'(D'D \otimes Q^{-1})U + \Sigma_0^{-1})^{-1}$, $c^* = \left(\Sigma^*(U' vec(Q^{-1}(X^*)'D) + \Sigma_0^{-1} c_0')\right)'$, where $U = [I_m \quad F]'$, $D = [1_n \quad -1_n]$. 1_n is the vector of ones with dimension $n \times 1$, while I_m is the identity matrix with $m \times m$ dimensions.

With the proper transformations of distributions (21)-(23) we can generate samples for parameters R, C, F and Q in each step of the Gibbs sampler for both regimes. The

selection of the starting covariance matrices V_0 and Σ_0 is crucial in a sense that it determines the parameter space explored.

We use the Random Walk Metropolis Hastings algorithm for the estimation of H. H_t is only dependent on $\lambda_{t,H}, \lambda_{t,E}, \lambda_{t,CDS}$. $\lambda_{t,k}$ is determined using the $\lambda_{t,k}^{new} = \lambda_{t,k} + \kappa_t \varepsilon_{t,k}$ equation, where $k \in \{H, E, CDS\}$, κ_t can be arbitrarily chosen and $\varepsilon_{t,k} \sim N(0,1)$ for every t and k. Using the generated parameter we calculate a candidate H, which we keep with a calculated α probability, after the evaluation of the measurement equation's likelihood function.

3.3 The Algorithm

Our parameter estimation algorithm is based on the works of Blake & Mumtaz [10]. The algorithm consists of the following 7 steps:

1. We choose appropriate starting points, which are in the parameter space. Our strategy is the following: we substitute the latent factors with their empirical ones. With these we estimate the system's transition equation using the OLS method. Based on the absolute error of the transition equation we organize and divide them into two groups. The empirical covariance matrices of these groups specify the starting covariance matrices. The Kim filter becomes applicable after the setting of the starting parameters.
2. Given the parameter space (Θ) of the model we run the Kim filter, thus we get the values of X_t for every time $t = 1, \dots, n$.
3. In the third step we use the $P(S_t = j|\mathcal{F}_t)$ probabilities acquired from the Kim filter to generate the probabilities of being in a given regime at time t starting from $t = n$ heading towards $t = 1$ using the transition probabilities. We determine the values of S_t for every $t = 1, \dots, n$ using random generation.
4. In the fourth step we generate new transition probabilities. Let π (1×2) be the vector of steady state probabilities, and P_{prior} the matrix of transition probabilities, then

$$M_{prior} = P_{prior} \circ \left(1_2 \otimes (\pi n)\right)$$

defines the prior state-transition matrix, with $\circ$ denoting the element-wise multiplication and $\otimes$ the Kronecker product. The empirical state-transition matrix M_{emp} comes from the realizations of S_t generated in step 3. Thus

$$P_i^{new} = Beta\left(M_{i,emp} + M_{i,prior}\right)$$

with i denoting the rows of the specified matrices ($i \in \{1,2\}$), while the Beta distribution comes as a special case of the Dirichlet distribution for $M = 2$ states[3].

[3] The general idea for the generation of transition probabilities follows the work of Kim and Nelson [7], whereas a detailed and intuitive description of the Dirichlet distribution is given by Frigyik, Kapila and Gupta [11].

5. Step five deals with the generation of the parameters Q, R, C, and F. We use random number generation based on the distributions suggested by Sugita [9] and Villani [8].
6. In step six we estimate H_t and $\lambda_{t,H}, \lambda_{t,E}, \lambda_{t,CDS}$ with the Random Walk Metropolis Hastings algorithm.
7. We repeat steps 2-6 until the convergence meets our expectations.

4 Data

In this section we give an overview of the macro-financial and yield curve data used for fitting the Markov regime switching model. We also address the challenges regarding the construction of yield curves and introduce our prior assumptions regarding the model. We use data with monthly frequency from October 2008 to December 2017 for a total of $n = 111$ observations for each time series. The limitation is due to the lack of earlier Hungarian CDS data. This gives us about 10 years of historical data and most importantly, includes the 2008 crisis and the following recession period, giving merit to using a regime switching model.

Our data sources include Thomson Reuters for CDS curve and Euro area yield curve points, the Hungarian Central Bureau of Statistics for Hungarian CPI data, Eurostat for the Euro area CPI data and the Hungarian National Bank for EUR/HUF exchange rate time series.

To obtain Hungarian yield curve data, we use the secondary market bid-ask quotes of HUF denominated Hungarian government bonds, as price quotes provided by primary dealers are considered as actual market prices, the fixing process of which is regulated by the Hungarian Government Debt Management Agency.

We use the Fama-Bliss [12] method to transform price quotes into yield curve points upon which we can fit the Nelson-Siegel model. This nonparametric method allows us to obtain zero-coupon yield curve points from bond price quotes. The Nelson-Siegel model can then be fitted on the unsmoothed Fama-Bliss yields.

We impose additional restrictions compared to the original Fama-Bliss [12] algorithm, using only fixed interest rate bonds with a maturity over 1 year as well as zero-coupon T-bills. Our implementation of the Fama-Bliss [12] algorithm can be summarized in four steps:

1. We determine whether the Yield-to-Maturity (YTM) of a given security is between the moving average of the previous 3 and the next 3 (based on time to maturity) government securities' yields.
2. We examine the unsmoothed Fama-Bliss forward yields. We exclude a government security from the analysis if the forward rate corresponding to its maturity shows a change greater than 0.2 percentage points in the opposite direction. In this step we calculate the forward rates using only government securities passing step 1.
3. We calculate the forward yield curve once more. Using the remaining government securities after step 2, we calculate the moving averages as in step 1. An instrument is to be considered in the analysis if the forward rate corresponding to its maturity is

between the two moving averages or the difference from one of them is not larger than 0.2 percentage points.

4. We repeat steps 2 and 3 using instruments excluded from the analysis but fulfilling the initial conditions.

In the end, we get a forward yield curve that fits the two smoothness criteria defined in steps 2 and 3. The forward yield curve can be then transformed into a zero-coupon yield curve.

Fitting a Markov regime switching model requires us to give an initial estimation of the transition probability matrix (Eq. 5). We set the starting value to:

$$P_{prior} = \begin{pmatrix} 0.987 & 0.013 \\ 0.180 & 0.820 \end{pmatrix}.$$

This means that the probability of going from the first regime to the second would be 1.3%, while exiting from the second regime would be 18%. Thus, the first regime is expected to last 6.5 years, whereas the expected duration of the second regime is only half a year. The prior steady state values of the factors were determined using P_{prior} and incorporate the effects of the current historically low interest rate period.

The Hungarian CPI expectation in the first regime equals the inflation target of the Hungarian Central Bank ($cpi_{prior,HUN} = 3\%$), and is one percentage point higher in the second regime. We expect the inflation rate of the Euro area to remain below the target set by the European Central Bank ($cpi_{prior,EU} = 1.5\%$). Regarding the HUF yield curve, we expect a 50bp increase in the short rate compared to 2017 December, and a 3.5% level for the long end of the yield curve. Our prognosis for the yield curve also includes a flattening slope and a rising level in case of a crisis.

Finally, we expect the EUR/HUF exchange rate to stagnate during the first regime and rise during the second regime, meaning a comparatively weaker HUF and stronger Euro in a crisis situation.

5 Results

This section is split into two parts. First, we give an overview of the results of the fitted model, including regime probabilities and steady state values. Finally, we make a 5-year forecast from end-2017 to end-2022 with 1 000 scenarios and examine its results[4].

5.1 Results of Model Fitting

The resulting transition probability matrix is somewhat different compared to the initial values:

$$P_{posterior} = \begin{pmatrix} 0.95 & 0.05 \\ 0.20 & 0.80 \end{pmatrix}.$$

[4] The length of the forecast aligns with the medium term debt management framework of the Hungarian Government Debt Management Agency.

According to the posterior probabilities, the normal regime lasts for two years on average while a typical crisis lasts for 5 months.

Regime	L_H	S_H	C_H	L_E	S_E	C_E	cpi_H	cpi_E	EUR/HUF	L_{CDS}	S_{CDS}
Normal	0.40	3.22	-3.10	-0.17	1.21	-1.15	2.79	1.36	-0.003	0.35	-2.64
Crisis	2.92	1.31	-0.10	0.28	1.03	-0.66	3.96	1.58	-0.011	0.97	-1.53

Table 1. Steady state factor values by regime

Table 1 shows the resulting steady state values. The Hungarian short rate is 0.4% in the normal regime and about 3% during a crisis, while the long rates are 3.6% and 4.2%, respectively. Steady state values for Hungarian inflation are 2.8% in the normal and 4% in the crisis regime.

Euro area short rates are close to zero in both regimes, while long rate steady state values are over 1%. Euro area inflation rates are consistent with our assumptions and do not reach the 2% ECB goal in any regime. The EUR/HUF exchange rate steady state values show the HUF getting slightly weaker compared to the Euro in the normal regime and weaker during a crisis.

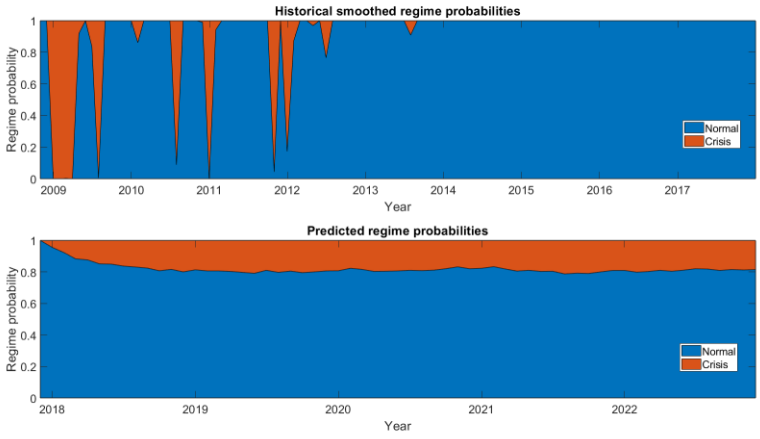


Fig. 1. Historical (upper panel) and predicted (lower panel) regime probabilities

Figure 1 shows the historical and predicted regime probabilities. Historical crisis occurrences can be easily connected to the recession period and the Greek debt crisis. The Hungarian economy has been relatively stable since 2014, which is also shown by the dominance of the normal regime since then. We can also see that from 2018 onwards we have an approximately 20% probability of slipping into the crisis regime, which is consistent with the estimated transition probabilities.

Although our experience suggests that using a regime switching model with the parameters described in this paper improves the robustness of the model, a quantitative proof is out of the scope of this study.

We have used 75 000 runs for the Gibbs sampling method. We discard the first 20 000 runs to alleviate the dependence on manually given prior values. Geweke test statistics [13] show that the simulation converges and 75 000 runs are enough for sampling.

5.2 Forecast Results

In order to make our predictions, we substitute the parameters gained from fitting the model into the transition equation. We generate 1 000 scenarios, all consisting of the trajectories of the yield curve and other macro-financial factors. The forecast horizon is 5 years with a monthly frequency.

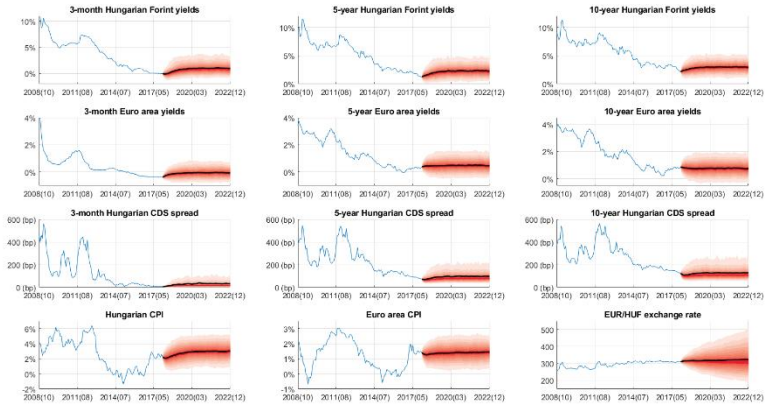


Fig. 2. Fan charts of selected yields, CDS spreads, CPI rates and exchange rate with 5-95% confidence intervals

Figure 2 shows the historical and predicted values of several yield curve points, CDS spreads and macro variables. We can see a gradual increase in Hungarian yield levels. By 2020, we predict a 1% short rate and a 3% 10-year yield. According to our forecast, the Euro area yields are mostly expected to stagnate with the short rate slowly converging to zero. Hungarian CDS spreads are expected to rise by $10-20$ basis points. The HUF shows a slight depreciation compared to the Euro. The fan charts also show that risks leading to the depreciation of the HUF as well as risks of higher interest rates are significantly higher than vice versa. CDS spreads in particular are especially vulnerable and can top 500bps in a crisis situation.

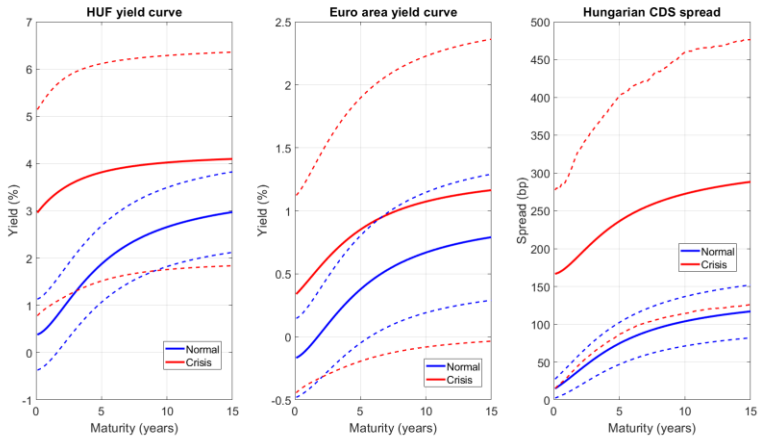


Fig. 3. Comparison of normal and crisis regime yield curves and CDS curves

Figure 3 shows that the two regimes are significantly different. The dashed lines show the standard deviation for the yield curves and the 16th and 84th percentile for the Hungarian CDS spread. They indicate a higher instability during the crisis regime for the CDS spread and both yield curves. The uncertainty of the Euro area yield curve is lower than the HUF yield curve during both regimes showing the inherent vulnerability of a small open economy compared to a large one. Yield curve and CDS levels are generally higher in the crisis regime. The Hungarian yield curve is also distinctly flatter in the second regime.

6 Conclusion

In this paper we used a Markov regime switching state-space model to forecast the Hungarian and Euro area yield curves with additional macro-financial factors on a 5-year horizon. We used a rotated Nelson-Siegel factor model for yield curve modeling and implemented the Kim filter to estimate the unobservable parameters. The fitted model matches our economic assumptions regarding normal and crisis regimes. We used data from October 2008 to December 2017. As of May 2018, the actual evolution of the factors is well within the confidence bounds of our predictions.

Further studies can be conducted to improve both the depth and the breadth of this analysis. A deeper analysis could include several competing models. Comparing and testing them using out-of-sample tests could highlight the different strengths of various models. A broader analysis could use the results of the forecast to predict the cash-flows of Hungarian government debt securities and calculate the costs of these together with the distribution of the costs, which is one of the risk factors that public debt managers face. Thus, a cost-risk optimization or a scenario analysis can be carried out on the Hungarian debt portfolio.

References

1. Nelson, C.R., Siegel, A.F.: Parsimonious Modeling of Yield Curves. The Journal of Business. 60, 473–489 (1987).
2. Diebold, F.X., Rudebusch, G.D.: Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. Princeton University Press (2013).
3. Nyholm, K.: A rotated Dynamic Nelson-Siegel model with macro-financial applications. European Central Bank (2015).
4. Kalman, R.: A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME - Journal of basic Engineering. 82, 35–45 (1960).
5. Hamilton, J.D.: A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica. 57, 357–384 (1989).
6. Kim, C.-J.: Dynamic linear models with Markov-switching. Journal of Econometrics. 60, 1–22 (1994).
7. Kim, C.-J., Nelson, C.R.: State-space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications. MIT Press (1999).
8. Villani, M.: Steady-state priors for vector autoregressions. Journal of Applied Econometrics. 24, 630–650 (2009).
9. Sugita, K.: Bayesian analysis of a vector autoregressive model with multiple structural breaks. Economics Bulletin. 3, 1–7 (2008).
10. Blake, A., Mumtaz, H.: Applied Bayesian econometrics for central bankers. Centre for Central Banking Studies, Bank of England (2012).
11. Frigyik, B., Kapila, A., Gupta, M.: Introduction to the Dirichlet Distribution and Related Processes. University of Washington (2010).
12. Fama, E.F., Bliss, R.R.: The Information in Long-Maturity Forward Rates. The American Economic Review. 77, 680–692 (1987).
13. Geweke, J.F.: Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Federal Reserve Bank of Minneapolis (1991).

Ranking multi-step system forecasts invariant to linear transformations

Håvard Hungnes

September 4, 2018

Abstract

The paper derives a test for equal predictability of multi-step system forecasts invariant to linear transformations. The derived test is a multivariate version of the Diebold-Mariano test. An invariant measure for forecast accuracy is necessary as the conclusions otherwise can depend on how the forecasts are reported (e.g., as in level or growth rates). The test is used in comparing quarterly multi-step system forecasts made by Statistics Norway with similar forecasts made by Norges Bank.

Keywords: Macroeconomic forecasts; Econometric models; Forecast performance; Forecast evaluation; Forecast comparison.

JEL codes: C32, C53.

Acknowledgments: Thanks to the referees at ITISE 2018 conference for valuable comments on an earlier version of this paper.

1 Introduction

Clements and Hendry (1993) showed that evaluation of forecasts of individual variables at each horizon separately is not invariant to linear transformations of the forecasts. Ericsson (2008) illustrates this by considering two different models for forecasting the oil price, where the multi-step forecasts based on one of the models are best when considering the oil price in levels, but that the forecasts of another model are best when considering the oil price growth. Clements and Hendry (1993) suggested a measure of the whole system of forecasts when evaluating the system forecasts. However, 25 years later Hendry and Martinez (2017) point out that "relatively little work has been done to evaluate the accuracy of the whole system jointly."

Usually, measures for forecast accuracy only consider forecast for one variable at one forecasting horizon. Measures such as mean absolute forecast errors and mean square forecast errors (or variants of these) are usually applied. For measuring the accuracy of a system of forecasts variants of these individual measures can be applied. One example is the mean (or sum) of the mean square forecast errors. Unfortunately, none of these measures are invariant to linear transformations of the forecasts (such as measuring the forecast errors in growth rates instead of log-levels). However, as suggested by Clements and Hendry (1993), using the determinant of the covariance matrix of the forecast errors of the whole system of forecasts yields a measure that is invariant to scale-preserving linear transformations of the forecasts. This measure is equivalent to the predictive likelihood, see Bjørnstad (1990).

Hendry and Martinez (2017) mention almost 20 papers that apply the determinant of the covariance matrix to evaluate system forecasts. However, these papers only rank sets of forecasts based on this measure and do not apply any formal test for testing if one system forecast is statistically better than another system forecast or if one system forecast encompasses another system forecast.

There are some important contributions in evaluating system forecasts. Jordà and Marcellino (2010) consider the forecasts of a variable over all forecasting horizons and derive the confidence regions of this path based on the covariance matrix of the forecast errors. By doing so, the confidence region will be independent of linear transformations of the forecasts. Furthermore, considering systems of more variables, Jordà and Marcellino (2010) show how these confidence regions for the forecast path of one variable changes with different assumptions of the future path of other variables.

Sinclair et al. (2012, 2015) present a test for the absence of a forecast bias, for the full system forecast. Sinclair and Stekler (2013) apply this approach to test for biases in the revision of forecasts. Also, this test is independent of linear transformations of the forecasts.

With many variables and many forecasting horizons, the covariance matrix can become very large. Building on Abadir et al. (2014), Hendry and Martinez (2017) propose an estimator of the large covariance matrix that can be derived with relatively few forecast error observations for evaluating such forecasts systems.

This paper is mostly related to Diebold and Mariano (1995), who suggests a method for comparing two forecasts. They suggest a test of equal forecast ability based on the deviation of the mean square error of two univariate forecasts. The approach assumes a quadratic loss function. A rejection of the hypothesis of equal predictability implies that the forecast with the smallest mean square error is significantly better than the other forecast. The method has been extended to system forecasts, see Pesaran and Skouras (2002). However, in this system version, a weighting matrix must be applied. The present paper motivates using the covariance matrix of the optimally weighted forecast errors of the two forecast systems as this weighting matrix.

When making forecasts of the Norwegian economy, Statistics Norway applies an econometric model.

However, in the process of making the forecasts, the forecasts are also based on judgemental forecasts where other information than what is included in the econometric model is used to improve the forecasts (see also Lawrence et al., 2006, for a review of judgmental forecasting techniques). Hence, the forecasts we are evaluating are considered to be model-free forecasts. This implies that testing the forecast we are conducting here is not a part of testing the underlying model where we have to take into account estimation uncertainty. For an overview of evaluating models based on their forecasts, see Clark and McCracken (2013).

In this paper, we consider a test for testing if one set of forecast is significantly better than another set of forecasts. The tests are used to compare forecasts made by Statistics Norway with forecast made by Norges Bank. In the analysis, we investigate the forecasts of GDP, CPI and the unemployment rate for the current and the next year jointly.

The rest of the paper is organized as follows: In Section 2 the theoretical background for the test as well as the proposed test for equal predictability is presented. In Section 3 the proposed ranking test is applied to compare the forecasts made by Statistics Norway with forecasts made by Norges Bank. Section 4 concludes.

2 Theory

2.1 Measures of forecast accuracy

Let $y^i_{t+h|t}$ be the forecast of variable i in period $t+h$ made in period t. In the present paper we assume that the value of y in period t is not known in period t; hence forecasts for the current period — also referred to as nowcasting — can be made and is denoted $y^i_{t|t}$. The prediction error of the forecast of variable i in period $t+h$ made in period t is defined as

$$e^i_{t+h|t} \equiv y^i_{t+h} - y^i_{t+h|t}, \tag{1}$$

where y^i_{t+h} is the outcome of variable i in period $t+h$.

The observed Mean Square Forecast Error (MSFE) is given by

$$T^{-1} \sum_{t=1}^{T} \left(e^i_{t+h|t} \right)^2, \tag{2}$$

which expresses the mean square forecast error of variable i forecasted h periods for forecasts made in T consecutive periods. The MSFE (or the root of MSFE) is a widely used measure for the accuracy of forecast and comparing of forecasts. However, it can be problematic to use this measure when comparing a system forecast of multiple variables or multiple forecasting horizons. To see the former, consider the following example:

Example 1 *Consider a nowcasts of a system with two variables; private consumption and income, both measured in logs. Assume a forecaster predicts both income and consumption to increase by 2 percent. Then, the National Accounts numbers shows that both variables increased by 3 percent that period, which implies that this forecaster missed by one percentage point for both private consumption and income. However, we could also consider the forecast accuracy of the system forecast by measuring the system MSFE of private consumption and private savings ratio, where the latter is the difference between (log of) private income and (log of) private consumption. Then this forecaster still would miss 1 percentage point on private income but be spot on for the forecast on the*

savings ratio.

Suppose, as in the example, that variable 1 is (log of) consumption and variable 2 is (log of) income. One measure of the accuracy of nowcasts of these two variables could be the sum of the MSEF of the two variables,

$$T^{-1}\sum_{t=1}^{T}\left(e^1_{t|t}\right)^2+T^{-1}\sum_{t=1}^{T}\left(e^2_{t|t}\right)^2. \tag{3}$$

However, we could alternatively consider the sum of the MSFE errors of (log) consumption and the savings ratio (defined as the difference between the log of income and the log of consumption). This measure would then be

$$T^{-1}\sum_{t=1}^{T}\left(e^1_{t|t}\right)^2+T^{-1}\sum_{t=1}^{T}\left(e^2_{t|t}-e^1_{t|t}\right)^2, \tag{4}$$

which is not identical to the previous measure.

To make a measure that is invariant of linear transformation of the variables within one forecasting horizon, we define a vector of the forecast of all N variables in period $t+h$ made in period t as $\mathbf{y}_{t+h|t}=\left(y^1_{t+h|t},y^2_{t+h|t},\ldots,y^N_{t+h|t}\right)'$. Similarly, the outcome of these variables in period $t+h$ is $\mathbf{y}_{t+h}=\left(y^1_{t+h},y^2_{t+h},\ldots,y^N_{t+h}\right)'$, which implies that the prediction error vector becomes $\mathbf{e}_{t+h|t}=\left(e^1_{t+h|t},e^2_{t+h|t},\ldots,e^N_{t+h|t}\right)'$ with elements defined as in (1). A matrix version of the observable MSFE for forecasting horizon h would then be

$$\hat{V}_h=T^{-1}\sum_{t=1}^{T}\mathbf{e}_{t+h|t}\mathbf{e}^{i\prime}_{t+h|t}. \tag{5}$$

This matrix is of dimension $N\times N$ and it is not obvious how to compare forecasts based on this measure. The measure that we considered above and showed was not invariant to linear transformation, corresponds to use the trace of the matrix in (5): if $N=2$ and $h=0$, than the trace is given by (3).

To compare the vectors of prediction errors we define $(V_h)^A$ and $(V_h)^B$ of two different forecasting models. Furthermore, let $\mathbf{M}(V_h)^A\mathbf{M}'$ and $\mathbf{M}(V_h)^B\mathbf{M}'$ be the similar measure of the linear transformed versions of the two forecasts where $\mathbf{M}$ is an $N\times N$ full rank matrix expressing the linear transformation.[1] For example, if (as in Example 1) the the first element in $\mathbf{e}_{t+h|t}$ refers to (the log of) private consumption an the latter refers to (the log of) private income, then pre-multiplying the vector prediction errors with $\mathbf{M}=\begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}$ transforms the vector of prediction errors to a vector where the first element still represents the prediction error for (the log of) private consumption but where the second element now is the prediction error of the savings ratio. Now the trace of (5) with these transformed forecasts are given by (4). Therefore, if $trace\left((V_h)^A\right)<trace\left((V_h)^B\right)$, it does not follow that $trace\left(\mathbf{M}(V_h)^A\mathbf{M}'\right)<trace\left(\mathbf{M}(V_h)^B\mathbf{M}'\right)$.

Clements and Hendry (1993) suggest using the determinant of (5) when comparing system forecasts. If $|(\hat{\mathbf{V}}_h)^A|<|(\hat{\mathbf{V}}_h)^B|$, then the forecast based on model A with MSFE given by $(V^i_H)^A$ has a smaller prediction error than the forecast of model B with MSFE given by $(\hat{\mathbf{V}}_h)^B$. This measure is invariant to linear transformations of the forecasts, i.e. $|(\hat{\mathbf{V}}_h)^A|<|(\hat{\mathbf{V}}_h)^B|$ implies $|\mathbf{M}(\hat{\mathbf{V}}_h)^A\mathbf{M}'|<|\mathbf{M}(\hat{\mathbf{V}}_h)^B\mathbf{M}'|$.[2] Furthermore, if $(\hat{\mathbf{V}}_h)^B-(\hat{\mathbf{V}}_h)^A\succ 0$, (i.e., that the difference between the two matrices defined by (5) is positive definite, see, e.g. Dhrymes, 1984, prop. 66), then it follows that $|(\hat{\mathbf{V}}_h)^A|<|(\hat{\mathbf{V}}_h)^B|$.

The MSFE (or the root of MSFE) is a widely used measure for the accuracy of forecast also for $h>0$;

[1] Clements and Hendry (1993) also assume $|\mathbf{M}|=1$, which can be interpreted as a scale-preserving transformation. Then $trace(\mathbf{M}(V_h)^A\mathbf{M}')\neq trace((V^i_H(V_h)^A)$ shows the lack of invariance. However, since two forecasts are compared, we do not need $\mathbf{M}$ to have this property; see also Schmidt (1993).

[2] The implication follows from $|\mathbf{M}\,(\hat{\mathbf{V}}_h)^j\mathbf{M}'|=|\mathbf{M}|^2|(\hat{V}_h)^j|$ for $j=A,B$.

see, e.g., Bjørnland et al. (2017), El-Shagi et al. (2016), Jungmittag (2016), and Kock and Teräsvirta (2016) for some recent applications. However, the MSFE for measuring the forecast accuracy when $h > 0$ depends on how the forecasts are measured, see Clements and Hendry (1993). Only in the case with one variable and $h = 0$ (univariate nowcasting) comparison based on the observed (univariate) MSFE are invariant of linear transformations of the forecasts, see Clements and Hendry (1993, 1998).

To compare forecasts generated by different models, we need to consider all forecasts up to forecast horizon H (where H is used for the longest forecast horizon). Therefore, we define $\mathbf{Y}_{t,H|t}$ to be the vector of forecasts of $\mathbf{y}_{t+h|t}$ in each period from period t to period $t+H$ made at time t, i.e., $\mathbf{Y}_{t,H|t} = \left(\mathbf{y}_{t|t}, \mathbf{y}_{t+1|t}, \ldots, \mathbf{y}_{t+H|t}\right)'$. The prediction error of $\mathbf{Y}_{t,H|t}$ is given by $\mathbf{E}_{t,H|t} \equiv \mathbf{Y}_{t,H} - \mathbf{Y}_{t,H|t}$, where $\mathbf{Y}_{t,H} = (\mathbf{y}_t, \mathbf{y}_{t+1}, \ldots, \mathbf{y}_{t+H})'$ is the vector of the outcome of all the variables from period t to period $t+H$. This implies that vector of prediction error is $\mathbf{E}_{t,H|t} = \left(\mathbf{e}_{t|t}, \mathbf{e}_{t+1|t}, \ldots, \mathbf{e}_{t+H|t}\right)'$.

A matrix version of the observable MSFE for all forecasting horizons up to period H would then be

$$\mathbf{V}_H = T^{-1} \sum_{t=1}^{T} \mathbf{E}_{t,H|t} \mathbf{E}'_{t,H|t}, \tag{6}$$

which is here denoted the MSFEM (for the MSFE Matrix).[3] This matrix is of dimension $K \times K$ with $K = N(H+1)$. Also for this measure, Clements and Hendry (1998) suggest using the determinant as a measure of accuracy.

2.2 Autocorrelation

Suppose the process of the considered variables can be formulated by the following Wold representation

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \Gamma_i v_{t-i},$$

with $\Gamma_0 = I_N$. Hence, the optimal h period ahead prediction given at time t and provided that the coefficients in the infinite lag polynomial $I_N + \Gamma_1 L + \Gamma_2 L^2 + \ldots$ are known and that the error in period t is not known, is

$$\mathbf{y}_{t+h|t} = \sum_{i=h+1}^{\infty} \Gamma_i v_{t+h-i},$$

and the prediction error then becomes

$$\mathbf{e}_{t+h|t} = \mathbf{y}_{t+h} - \mathbf{y}_{t+t|t} = \sum_{i=0}^{h} \Gamma_i v_{t+h-i}.$$

Hence,

$$\begin{pmatrix} \mathbf{e}_{t|t} \\ \mathbf{e}_{t+1|t} \\ \vdots \\ \mathbf{e}_{t+H|t} \end{pmatrix} = \begin{pmatrix} I_N & 0 & \cdots & 0 \\ \Gamma_1 & I_N & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \Gamma_H & \cdots & \cdots & I_N \end{pmatrix} \begin{pmatrix} v_t \\ v_{t+1} \\ \vdots \\ v_{t+H} \end{pmatrix},$$

which shows that optimal forecasts up to a forecasting horizon H has autocorrelation of order H. The reason is that a forecast made in period $t+H$ will overlap with a forecast made in period t, as both sets

[3] Clements and Hendry (1998) referrers to the measure in (6) as GFESM (generalized forecast error second moment) and the measure in (5) as MSFEM.

of forecasts will involve forecasts of variables for period $t+H$. However, a optimal forecast made in period $t+H+1$ will not overlap with a forecast made in period t, and hence the prediction errors in the two sets of forecasts are not expected to be correlated. This property is also shown in Hendry and Martinez (2017) and used by Harvey et al. (1997, 1998), and Harvey and Newbold (2000), among others.

2.3 The Diebold-Mariano test

Diebold and Mariano (1995) suggest a test for equal predictability. Let $e^{i,j}_{t+h|t}$ $(j = A, B)$ be the prediction error of the forecast made by forecaster j of variable i in periode $t+h$ made in period t. Then consider a loss-difference series $d^i_{t,h} = \left(e^{i,A}_{t+h|t}\right)^2 - \left(e^{i,B}_{t+h|t}\right)^2$ for variable i at forecasting horizon h when considering a quadratic loss function. The test statistics Diebold and Mariano (1995) suggests is simply

$$T^{1/2}\bar{d}^i_h Q^{-1/2}_{h,i}, \tag{7}$$

where $\bar{d}^i_h = T^{-1}\sum_{t=1}^{T} d^i_{t,h}$ (i.e., ...) and

$$Q_{h,i} = \frac{1}{T}\left[\sum_{t=1}^{T}\left(d^i_{t,h} - \bar{d}^i_h\right)^2 + 2\sum_{l=1}^{\tau}\sum_{t=1}^{T-l}\left(d^i_{t,h} - \bar{d}^i_h\right)\left(d^i_{t+l,h} - \bar{d}^i_h\right)\right], \tag{8}$$

where τ is the truncation lag, $\tau geq T$. Diebold and Mariano (1995) show that the test statistics in (7) is asymptotical standard normally distributed. The null hypothesis of $\bar{d}^i_h = 0$ implies equal predictability. If this null hypothesis is rejected and $\bar{d}^i_h < 0$, then forecast A of variable i at horizon h is significantly better than forecast B.

The test statistics in (7) only considers a univariate forecast, i.e., a forecast of one variable at one forecasting horizon. Pesaran and Skouras (2002) presents the loss-difference series

$$d_t = \mathbf{E}^{A\prime}_{t,H|t}{}'\mathbf{H}\mathbf{E}^{A}_{t,H|t} - \mathbf{E}^{B\prime}_{t,H|t}{}'\mathbf{H}\mathbf{E}^{B}_{t,H|t} \tag{9}$$

for the multivariate quadratic model where the $K \times K$ matrix $\mathbf{H}$ depends on the parameters in the loss function. Capistrán (2006) suggests using $\mathbf{H} = I_K$ and Quaedvlieg (2017) suggests (in his weighted average loss test) using a diagonal matrix with weights along its diagonal. However, none of these suggestions leads to a measurement that is invariant to linear transformations.

2.4 The test statistics for equal predictability

Williams and Kloot (1953) introduced a general test for testing equal predictability for two models, see also Howrey (1993). Consider two different forecasts of variable i in period $t+h$ made in period t; denoted $y^{i,A}_{t+h|t}$ and $y^{i,B}_{t+h|t}$. Then, we apply the regression

$$y^i_{t+h} - \frac{1}{2}\left[y^{i,A}_{t+h|t} + y^{i,B}_{t+h|t}\right] = \gamma\left[y^{i,B}_{t+h|t} - y^{i,A}_{t+h|t}\right] + u^i_{t+h|t}. \tag{10}$$

A test of the hypothesis $\gamma = 0$ is a test of equal weight on both forecasts. If the estimate of γ is negative, we can test if the coefficient is significant. If it is, we say that forecast A is significantly better than forecast B.

Note also that the hypothesis that $\gamma = -1/2$ corresponds to a of forecast A encompassing forecast B, i.e., forecast A contains all information so there is no additional information in forecast B. This test is

not considered here, see Hungnes (2018) for a system version of this test.

A generalization of this test could be to consider the vector version where all forecast horizons up to H for all variables are considered jointly, i.e.

$$\mathbf{Y}_{t,H} - \frac{1}{2}\left[\mathbf{Y}^A_{t,H|t} + \mathbf{Y}^B_{t,H|t}\right] = \gamma\left[\mathbf{Y}^B_{t,H|t} - \mathbf{Y}^A_{t,H|t}\right] + \mathbf{U}_{t,H|t}, \tag{11}$$

where the error term is $\mathbf{U}_{t,H|t} = \left(\mathbf{u}^{1\prime}_{t,H|t}, \mathbf{u}^{2\prime}_{t,H|t}, \ldots, \mathbf{u}^{N\prime}_{t,H|t}\right)'$ with $\mathbf{u}^i_{t,H|t} = \left(u^i_{t|t}, u^i_{t+1|t}, \ldots, u^i_{t+H|t}\right)'$. This can also be formulated as

$$y_t = \gamma x_t + \mathbf{U}_{t,H|t}, \tag{12}$$

where $y_t = \frac{1}{2}\mathbf{E}^A_{t,H|t} + \frac{1}{2}\mathbf{E}^B_{t,H|t}$ and $x_t = \mathbf{E}^A_{t,H|t} - \mathbf{E}^B_{t,H|t}$.

By defining the vectors $\mathbf{y} = \left(y'_1, y'_2, \ldots, y'_T\right)'$ and $\mathbf{x} = \left(x'_1, x'_2, \ldots, x'_T\right)'$, (12) can be formulated as

$$\mathbf{y} = \gamma\mathbf{x} + \varepsilon, \tag{13}$$

with $\varepsilon = \left(\mathbf{U}'_{1,H|1}, \mathbf{U}'_{2,H|2}, \ldots, \mathbf{U}'_{T,H|T}\right)'$.

The conditional estimators for γ and the covariance matrix for ε in (13) are given by (when ignoring possible degrees of freedom adjustments for the covariance matrix)

$$\hat{\gamma}_{(\hat{\Omega})} = \left(\frac{1}{T}\mathbf{x}'\left(I_T \otimes \hat{\Omega}^{-1}_{(\hat{\gamma})}\right)\mathbf{x}\right)^{-1}\left(\frac{1}{T}\mathbf{x}'\left(I_T \otimes \hat{\Omega}^{-1}_{(\hat{\gamma})}\right)\mathbf{y}\right), \tag{14}$$

$$\hat{\Omega}_{(\hat{\gamma})} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{\gamma}_{(\hat{\Omega})}x_t\right)\left(y_t - \hat{\gamma}_{(\hat{\Omega})}x_t\right)', \tag{15}$$

where $\otimes$ indicates the Kronecker product and the subscript in parenthesis indicates that the estimates are a function of another estimate. The estimates in (14) and (15) can be obtained by an iterative procedure until convergence, and the final estimates will equal the ones obtained with full information maximum likelihood, see Oberhofer and Kmenta (1974).

The estimator of the variance of (14), when considering that forecasts made at different time periods can be correlated, is given by

$$\begin{aligned}&\widehat{Var\left(\hat{\gamma}_{(\hat{\Omega})}\right)}_{(g)}\\ =&\frac{1}{T}\left[\frac{1}{T}\mathbf{x}'\left(I_T \otimes \hat{\Omega}^{-1}_{(\hat{\gamma})}\right)\mathbf{x}\right]^{-2}\left[\frac{1}{T}\mathbf{x}'\left(I_T \otimes \hat{\Omega}^{-1}_{(\hat{\gamma})}\right)\hat{\Sigma}_{(g)}\left(I_T \otimes \hat{\Omega}^{-1}_{(\hat{\gamma})}\right)\mathbf{x}\right],\end{aligned} \tag{16}$$

where the $TK \times TK$ matrix $\hat{\Sigma}_{(a)}$ is the covariance matrix of ε. The parameter g usually is set equal to the estimated γ or γ under the null hypothesis (here; 0). Section 2.2 shows that with optimal forecasts H steps ahead (including nowcasting), there will be autocorrelation up to order H and no autocorrelations above order H. Hence, the block element $(t, t+l)$ in $\hat{\Sigma}_{(a)}$ is given by

$$\{\hat{\Sigma}_{(g)}\}_{t,t+l} = \begin{cases}\hat{\mathbf{U}}_{(g),t,H|t}\hat{\mathbf{U}}'_{(g),t+l,H|t+l} & \text{for } l = -H, \ldots, 0, \ldots, H\\ \mathbf{0}_{K\times K} & \text{otherwise,}\end{cases} \tag{17}$$

where

$$\hat{\mathbf{U}}_{(g),t,H|t} = y_t - gx_t, \tag{18}$$

which implies that $\hat{\mathbf{U}}_{(0),t,H|t} = y_t$.

The last term in (16) can then be written as

$$\begin{aligned}\mathbf{Q}_{(g)} =& \frac{1}{T}\mathbf{x}'\left(I_T \otimes \hat{\Omega}_{(\hat{\gamma})}^{-1}\right)\hat{\Sigma}_{(g)}\left(I_T \otimes \hat{\Omega}_{(\gamma)}^{-1}\right)\mathbf{x} \\ =& \frac{1}{T}\left[\sum_{t=1}^{T} d_{(g),t}^2 + 2\sum_{l=1}^{H}\sum_{t=1}^{T-l} w_l d_{(g),t} d_{(g),t+l}\right],\end{aligned} \tag{19}$$

where

$$d_{(g),t} = x_t'\hat{\Omega}_{(\hat{\gamma})}^{-1}\hat{\mathbf{U}}_{(g),t,H|t}, \tag{20}$$

and where w_l must be equal to unity for the equality in (19) to hold. However, to secure that the estimated variance in the (16) is positive, Newey and West (1987) suggest using $w_l = 1 - \frac{l}{H+1}$ ($l = 1, \ldots, H$), as will be used here.

The expression for $d_{(g),t}$ in (20) resembles the multivariate version of Diebold and Mariano (1995) loss differential. By inserting the expressions for x_t and y_t we have

$$\begin{aligned}2d_{(0),t} &= \left(\mathbf{E}_{t,H|t}^{A} - \mathbf{E}_{t,H|t}^{B}\right)'\hat{\Omega}_{(\hat{\gamma})}^{-1}\left(\mathbf{E}_{t,H|t}^{A} + \mathbf{E}_{t,H|t}^{B}\right) \\ &= \mathbf{E}_{t,H|t}^{A\prime}\hat{\Omega}_{(\hat{\gamma})}^{-1}\mathbf{E}_{t,H|t}^{A} - \mathbf{E}_{t,H|t}^{B\prime}\hat{\Omega}_{(\hat{\gamma})}^{-1}\mathbf{E}_{t,H|t}^{B},\end{aligned} \tag{21}$$

which is the multivariate version of the Diebold and Mariano (1995) loss differential based on square forecast errors and with $\hat{\Omega}_{(\hat{\gamma})}^{-1}$ as the weighting matrix.

Based on (20), two alternative t-tests can be formulated for testing the null hypothesis of $\gamma = 0$: one where the estimated value of γ is used in the expression for the variance in (16) and another where the value under the null hypothesis is used. Both t-statistics can be formulated as

$$\frac{\hat{\gamma}}{\sqrt{\widehat{Var}\left(\hat{\gamma}\right)_{(g)}}} = T^{1/2} w_0^{1/2} \bar{d}_{(0)} \mathbf{Q}_{(g)}^{-1/2}, \tag{22}$$

where $\bar{d}_{(g)}$ is the sample mean of (20). When the test statistic is computed for the estimated γ in the expression of the variance, we have $g = \hat{\gamma}$; and when it is computed under the null, we have $g = 0$. Based on Harvey et al. (1997) we also include the correction factor $w_0 = T^{-1}\left[T - 1 - 2H + T^{-1}H(H+1)\right]$. This correction factor is because one in (19) divide by the sample size T instead of the number of autocovariances $T - l$ (where $l = 1, 2, \ldots, H$).

In the univariate case, i.e., when only forecast of one variable and not of a vector of forecasts, Harvey et al. (1998) and Harvey and Newbold (2000) show that both variants of the test that are similar to (22) have severe size distortions; the version with $g = \gamma$ over-rejects and the other under-rejects in small samples. Therefore Harvey et al. (1998) and Harvey and Newbold (2000) suggest a modification of the test where $\mathbf{Q}_{(g)}$ in (22) is replaced with

$$\mathbf{Q}_{(g)}^{*} = \frac{1}{T}\left[\sum_{t=1}^{T}\left(d_{(g),t} - \bar{d}_{(g)}\right)^2 + 2\sum_{l=1}^{H}\sum_{t=1}^{T-l} w_l\left(d_{(g),t} - \bar{d}_{(g)}\right)\left(d_{(g),t+l} - \bar{d}_{(g)}\right)\right]. \tag{23}$$

They show that this modified expression with g equal to γ under the null hypothesis, the corresponding t-test has only small size distortions in the univariate case. Furthermore, the t-statistic in (22) with $\mathbf{Q}_{(0)}^{*}$ corresponds to the multivariate version of the Diebold and Mariano (1995) test statistics.

We will now show that the test statistics in (22) is independent of linear transformations of the

forecasts. It follows from both (19) or (23) that $\mathbf{Q}_{(g)}$ and $\mathbf{Q}^*_{(g)}$ are unaltered of such linear transformations if $d_{(g),t}$ and $\bar{d}_{(g)}$ are unaltered by such linear transformations. Hence, it is sufficient to show that $d_{(g),t}$ (and therefore $\bar{d}_{(g)}$) is invariant of linear transformations of the forecasts. If we consider the linear transformation given by the $K \times K$ matrix $\mathbf{M}$ such that $\mathbf{M}\left(\mathbf{E}^A_{t,H|t} - \mathbf{E}^B_{t,H|t}\right)$ is the difference between the two system forecasts and $\mathbf{M}\hat{\mathbf{U}}_{(g),t,H|t}$ is the corresponding residual vector to (18) of the transformed forecasts. Hence, the transformed covariance matrix is $\mathbf{M}\hat{\Omega}_{(a)}\mathbf{M}'$. By substituting these terms into the definition of $d_{(g),t}$ in (20), it follows that the transformation matrix cancels out and we have shown that $d_{(g),t}$ is independent of linear transformations of the forecasts.

Harvey et al. (1997) showed that the test in (22) have a distribution close to a t-distribution with $T-1$ degrees of freedom in the univariate case. In the multivariate case of a system of forecasts with a vector of K forecasts, we apply the t-distribution with $T-K$ degrees of freedom. Simulation results in Hungnes (2018) indicate that the 95 percent quantile in the statistic in (22) increases with K, and that the adjustment based on $T-K$ performs better than using $T-1$. An additional argument for using the smaller number of degrees of freedom is that t-tests based on heteroscedasticity-consistent standard errors generally exhibit substantial size distortions, see, e.g., Andrews (1991), Andrews and Monahan (1992), and den Haan and Levin (1997). Usually, the distortions lead to over-rejecting. Applying $T-K$ degrees of freedom instead of $T-1$ degrees of freedom in the test can therefore also reduce distortion bias.

2.5 Size and power of the tests

Hungnes (2018) presents a Monte Carlo simulation and investigates the size and power of an encompassing test. The results from the simple encompassing test also apply to the test presented here. Hence, the important findings there are repeated here.

Regarding the size of the test, the most important result is that the size distortion is smaller when applying the distribution under the null with (23). Hence, this is the statistic we apply in our comparison in the next section.

Regarding the power of the test, the Monte Carlo example in Hungnes (2018) indicates that the power increases with the dimension K in addition to the sample size T.

3 Comparison with Norges Bank

Statistics Norway has, with a few exceptions, published forecasts every quarter for many variables for the year the forecast was made as well as the following year since the 1st quarter in 1990. Among these variables are Mainland GDP[4], CPI and the unemployment rate (UR).

Even though a quarterly model is used in generating the forecasts, the forecasts are only published in annual terms. Therefore, in testing the forecasts from Statistics Norway, we only consider forecasts of annual values.

Statistics Norway did not publish forecasts in the 2nd quarter of 1990 and 1991. In the analysis, we have set those forecasts equal to the forecasts made in the 1st quarter. Also in the 3rd quarter in 2013 Statistics Norway did not publish a forecast. Here, also, the forecast for the 3rd quarter is set equal to

[4]Mainland Norway consists of all domestic production activity except exploration of crude oil and natural gas, transport via pipelines and ocean transport. The term was revised as a part of the main revision of the national accounts in 2014. Before this, service activities incidental to oil and gas were also excluded from Mainland Norway.

Table 1: Comparison with Norges Bank

	Q1 (1990-2014)	Q2 (1993-2014)	Q3 (1990-2014)	Q4 (1992-2014)
Both horizons				
$\hat{\gamma}_{\hat{\Omega}(\hat{\gamma})}$	0.0117 (0.0801)	-0.0215 (0.1017)	-0.2502 (0.0651)	0.1348 (0.0791)
$H_0 : \gamma = 0$	(0.0877) 0.0178 [0.8960]	(0.1064) 0.0407 [0.8429]	(0.0765) 10.6882 [0.0067]**	(0.0869) 2.4086 [0.1402]
Nowcasting				
$\hat{\gamma}_{\hat{\Omega}(\hat{\gamma})}$	-0.1328 (0.1448)	-0.0221 (0.1608)	-0.2386 (0.0844)	0.2805 (0.1018)
$H_0 : \gamma = 0$	(0.1430) 0.8623 [0.3678]	(0.1646) 0.0180 [0.8948]	(0.1148) 4.3183 [0.0553]	(0.1314) 4.5578 [0.0460]*
GDP $H = 1$				
$\hat{\gamma}_{\hat{\Omega}(\hat{\gamma})}$	-0.3056 (0.3937)	-0.3209 (0.2678)	-0.7963 (0.2442)	-0.0814 (0.1264)
$H_0 : \gamma = 0$	(0.4322) 0.5002 [0.4896]	(0.2936) 1.1950 [0.2880]	(0.3726) 4.5665 [0.0484]*	(0.1459) 0.3116 [0.5829]
CPI $H = 1$				
$\hat{\gamma}_{\hat{\Omega}(\hat{\gamma})}$	-0.0590 (0.1359)	-0.1506 (0.1905)	-0.1464 (0.0792)	0.2830 (0.1295)
$H_0 : \gamma = 0$	(0.1462) 0.1630 [0.6918]	(0.1728) 0.7601 [0.3942]	(0.1001) 2.1412 [0.1628]	(0.1767) 2.5657 [0.1249]
UR $H = 1$				
$\hat{\gamma}_{\hat{\Omega}(\hat{\gamma})}$	0.0604 (0.1557)	0.0336 (0.2262)	0.0986 (0.1680)	0.2677 (0.2313)
$H_0 : \gamma = 0$	(0.1849) 0.1067 [0.7482]	(0.2438) 0.0190 [0.8917]	(0.1733) 0.3235 [0.5774]	(0.2612) 1.0500 [0.3177]

Note: Q1 – Q4 indicates the quarter (of the year) the forecast is made. For $\hat{\hat{\Omega}}_{(\hat{\Omega})}$ the estimated value and its standard errors (implicitly derived such that the t-value is given in (22) with $g = \gamma$) is reported. For the hypothesis tests the standard errors (derived under the null hypothesis $\gamma = 0$ with $\mathbf{Q}^*_{(0)}$ from (23)), F-value (squared t-value), and the corresponding p-value to the F-value are reported, in addition to asterisks, where one asterisk denote that the test statistics exceeds the 95 percent critical value and two asterisks denote that the test statistics exceeds the 99 percent critical value. For the full test the F-value with corresponding p-value is reported together with asterisks.

the previously published forecast, i.e., the forecast from the 2nd quarter that year.[5] The latest forecast considered in the examination is the forecast made in the 4th quarter of 2014, which includes forecasts of variables for the year 2015. Hence, the sample spans 25 years of forecasts.

The published numbers for CPI and the unemployment rate are never revised. The published numbers for variables from the National Accounts, such as the Mainland GDP, can be revised in many quarters until they are fixed. Though also these 'fixed' numbers can be revised due to revisions in the System of National Accounts. In the analysis undertaken here, the first published number of Mainland GDP-growth is used.

Table 1 compares the forecasts made by Statistics Norway with the forecasts made by Norges Bank (the central bank of Norway). Forecasts from Norges Bank made in the 1st quarter of the year is available from 1996; and forecasts from Norges Bank made in the 2nd quarter are available from 1993. These forecasts are published about the same time as the forecasts from Statistics Norway are published. Hence, these forecasts can (without problems) be compared.

From 2001 to 2012 Norges Bank only published their forecasts 3 times a year. The last forecasts in these 12 years were published at the end of October. When comparing the forecast from the third quarter from Statistics Norway and Norges Bank, we use the forecasts made in the beginning of September for Statistics Norway. For Norges Bank, we use forecasts made about the same time in the years 1996-2000 and 2013-2014. For the years 2001-2012 we use the forecasts published in end-October. This implies that Norges Bank have about $1\frac{1}{2}$ month of information advantages in 12 out of 19 years we compare forecasts from the "3rd" quarter.[6]

Norges Bank has published forecasts in the 4th quarter since 1992. In the years 2001-2012 these forecasts were made in late October, while in the other years they are from December. The forecasts

[5] Due to the onset of the financial crises, Statistics Norway published an extra forecast in mid-October 2008. This extra forecast is not included in the current analysis.

[6] This differs form Bjørnland et al. (2012) who use the forecasts from the 2nd quarter when they compare the forecasts from Norges Bank with a system of averaging models, SAM.

from Statistics Norway are from beginning of December in all the years we compare. Hence, here Statistics Norway has an information advantages in 12 out of 23 years.

In the upper part of the table ('Both horizons') we compare the forecast for all three variables in both the current and the next year. In the first line, two numbers are reported for the forecasts made in each quarter; the estimated γ and its standard error.[7] For forecasts made in the 1st quarter (of the year) this estimate is about 0.01, indicating that the forecasts made by Statistics Norway is slightly better than the forecasts made by Norges Bank. However, seen in relation to the relative high standard error (0.08), this indicates that the estimate is not significantly different from zero.

In the next line, we indirectly apply (22) under the null hypothesis ($g = 0$) with $\mathbf{Q}^*_{(0)}$ from (23) to derive the standard error (reported in parenthesis). This leads to a slightly higher estimate of the standard error. The next two figures reported in the cell is the F-value of the hypothesis test (i.e., the squared t-value) and the corresponding p-value in brackets. As can be seen, we cannot reject the null hypothesis that the forecasts made by Statistics Norway and Norges Bank are equally good.

When comparing the forecasts made by Statistics Norway and Norges Bank made in the 2nd quarter of the year, we find that the estimate is not significantly different from zero (an estimate of -0.02 with a standard deviation given by 0.10). Hence, also for the forecasts made in the 2nd quarter made by Statistics Norway and Norges Bank we cannot reject that they are equally good.

For the forecasts made in the 3rd quarter, the estimated γ is close to $\frac{1}{4}$ and — by both measures of the standard error — is clearly significantly different from zero. This estimate imply that we could compute an optimal forecast with a weight of $\frac{3}{4}$ for the forecast made by Norges Bank and $\frac{1}{4}$ for the forecast made by Statistics Norway in this quarter. However, here Norges Bank has an information advantage of about 1,5 month in more than half of the years the forecasts were made. Hence, we would expect Norges Bank to do better than Statistics Norway for forecasts made in the 3rd quarter of the year.

For the forecasts made in the 4th quarter, the estimated γ is about 0.135, indicating that the forecasts made by Statistics Norway in this quarter are better than the forecasts made by Norges Bank. However, the forecasts made by Statistics Norway are not significantly better than the ones made by Norges Bank, despite that Statistics Norway has had an information advantage for many of the years.

In the next part of the table ('Nowcasting') we only consider the forecast for the current year for the three variables. The results for the forecasts made in the 1st and 2nd quarter are is similar to the forecast for both horizons. When only considering now-casting we do not find that the forecasts made by Norges Bank in the 3rd quarter is significantly better than the forecasts made by Statistics Norway, even though the point estimate is almost identical to the one when considering both forecasting horizons. For now-casts made in the 4th we see that the forecasts made by Statistics Norway are significantly better than the ones from Norges Bank. However, this must be contributed to that Statistics Norway has an information advantage in half of the years and this information advantages is much more important when only forecasts for the current year is considered.

In the bottom of the table the forecasts for Mainland-GDP, CPI, and the unemployment rate are considered separately to see which institution does the best to forecast these variables. For Mainland-GDP Norges Bank makes the best forecasts, based on the sign of the estimated γ. But only for Mainland-GDP are the forecasts by Norges Bank made in the 3rd quarter significantly better than the forecasts made by Statistics Norway.

[7] The standard error is implicitly derived such that the t-value is given in (22) with $g = \gamma$.

4 Conclusions

This paper presents a test for equal predictability of system forecasts. The test, which is a variant of a multivariate version of the Diebold and Mariano (1995) test, is invariant to linear transformations of the system forecasts. The test is used to compare the forecasts made by Statistics Norway with forecasts made by Norges Bank. We find that when the forecasts are made approximately at the same time, they are equally good in the sense that none of the are significantly better than the other. However, when one institution has an information advantage, the forecast made by the institution with an information advantage can be significantly better than the forecasts made by the other institution.

In the present paper we consider testing equal predictability between only two system forecasts. Mariano and Preve (2012) considers a test considers an extension of of the Diebold and Mariano (1995) test to test equal predictability between three or more univariate forecasts. Mariano and Preve (2012) show that this test is invariant concerning the ordering of the forecasts being compared. The test proposed by Mariano and Preve (2012) can easily be applied to compare multiple system forecasts when each pair of system forecasts is compared as in the present paper.

References

Abadir, K. M., Distaso, W., and Žikeš, F. (2014). Design-free estimation of variance matrices. *Journal of Econometrics*, 181(2):165–180.

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, 59(3):817–858.

Andrews, D. W. K. and Monahan, J. C. (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica*, 60(4):953–966.

Bjørnland, H. C., Gerdrup, K. R., Jore, A. S., Smith, C., and Thorsrud, L. A. (2012). Does Forecast Combination Improve Norges Bank Inflation Forecasts? *Oxford Bulletin of Economics and Statistics*, 74(2):163–179.

Bjørnland, H. C., Ravazzolo, F., and Thorsrud, L. A. (2017). Forecasting GDP with global components: This time is different. *International Journal of Forecasting*, 33(1):153–173.

Bjørnstad, J. F. (1990). Predictive Likelihood: A Review. *Statistical Science*, 5(2):242–254.

Capistrán, C. (2006). On comparing multi-horizon forecasts. *Economics Letters*, 93(2):176–181.

Clark, T. and McCracken, M. (2013). Advances in Forecast Evaluation. In *Handbook of Economic Forecasting*, volume 2, pages 1107–1201.

Clements, M. P. and Hendry, D. F. (1993). On the limitations of comparing mean square forecast errors. *Journal of Forecasting*, 12(8):617–637.

Clements, M. P. and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge University Press.

den Haan, W. J. and Levin, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Handbook of Statistics*, volume 15, chapter 12, pages 299–342. Elsevier.

Dhrymes, P. J. (1984). *Mathematics for Econometrics*. Springer-Verlag, New York, 2nd edition.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

El-Shagi, M., Giesen, S., and Jung, A. (2016). Revisiting the relative forecast performances of Fed staff and private forecasters: A dynamic approach. *International Journal of Forecasting*, 32(2):313–323.

Ericsson, N. R. (2008). Comment on 'Economic Forecasting in a Changing World' (by Michael Clements and David Hendry). *Capitalism and Society*, 3(2):1–16.

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1998). Tests for Forecast Encompassing. *Journal of Business & Economic Statistics*, 16(2):254–259.

Harvey, D. I. and Newbold, P. (2000). Tests for multiple forecast encompassing. *Journal of Applied Econometrics*, 15(5):471–482.

Hendry, D. F. and Martinez, A. B. (2017). Evaluating Multi-Step System Forecasts with Relatively Few Forecast-Error Observations. *International Journal of Forecasting*, 33(784):359–372.

Howrey, P. E. (1993). On the limitations of comparing mean square forecast errors: Comment. *Journal of Forecasting*, 12(8):652–654.

Hungnes, H. (2018). Encompassing tests for evaluating multi-step system forecast invariant to linear transformations.

Jordà, Ò. and Marcellino, M. (2010). Path forecast evaluation. *Journal of Applied Econometrics*, 25(4):635–662.

Jungmittag, A. (2016). Combination of Forecasts across Estimation Windows: An Application to Air Travel Demand. *Journal of Forecasting*, 35(4):373–380.

Kock, A. and Teräsvirta, T. (2016). Forecasting Macroeconomic Variables using Neural Network Models and Three Automated Model Selection Techniques. *Econometric Reviews*, 35(8-10):1753–1779.

Lawrence, M., Goodwin, P., O'Connor, M., and Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3):493–518.

Mariano, R. S. and Preve, D. (2012). Statistical tests for multiple forecast comparison. *Journal of Econometrics*, 169(1):123–130.

Newey, W. K. and West, K. D. (1987). A simple positive semi-definite heteroskedasticity and autocorrelation-consistent covariance matrix. *Econometrica*, 55:703–708.

Oberhofer, W. and Kmenta, J. (1974). A general procedure for obtaining maximum likelihood estimates in generalized regression models. *Econometrica*, 42(3):579–590.

Pesaran, M. H. and Skouras, S. (2002). Decision-Based Methods for Forecast Evaluation. In Clements, M. P. and Hendry, D. F., editors, *A Companion to Economic Forecasting*, pages 241–267.

Quaedvlieg, R. (2017). Multi-Horizon Forecast Comparison. *Working Paper*.

Schmidt, P. (1993). On the limitations of comparing mean square forecast errors: Comment. *Journal of Forecasting*, 12(8):660–662.

Sinclair, T. M. and Stekler, H. O. (2013). Examining the quality of early GDP component estimates. *International Journal of Forecasting*, 29(4):736–750.

Sinclair, T. M., Stekler, H. O., and Carnow, W. (2012). A new approach for evaluating economic forecasts. *Economics Bulletin*, 32(3):2332–2342.

Sinclair, T. M., Stekler, H. O., and Carnow, W. (2015). Evaluating a vector of the Fed's forecasts. *International Journal of Forecasting*, 31(1):157–164.

Williams, E. J. and Kloot, N. H. (1953). Interpolation in a series of correlated observations. *Australian Journal of Applied Science*, 4:1–17.

Conditional Heteroskedasticity in Long Memory Model 'FIMACH' for Return Volatilities in Equity Markets

Sabur Mollah[1]

A.M.M. Shahiduzzaman Quoreshi[2]

Abstract:

This paper incorporates conditional heteroscedasticity properties in the long memory model and applies the model on squared returns of BRICS (Brazil, Russia, India, China, and South Africa), UK and USA equity markets to capture the volatility of stock return. The conditional first- and second-order moments are provided. The CLS, FGLS and QML are discussed and 2SQML estimator is proposed. The simulation study suggests that the proposed 2SQML estimator performs better than the other three estimators. Both in simulation and empirical studies, we find that the proposed model FIMACH outperforms FIGARCH in terms of eliminating serial correlations.

Key Words: Long Memory Conditional Heteroskedastic Model, Return Volatility.

JEL Classification: C13, C22, C25, C51, G01, G12, G14, G17.

[1] Sabur Mollah, School of Management, Swansea University. E-mail: sabur.mollah@swansea.ac.uk

[2] Corresponding author. Blekinge Institute of Technology, SE-371 79 Karlskrona, Sweden. E-mail: shahiduzzaman.quoreshi@bth.se; Tel:+46734223619

1. Introduction

The volatility of stock returns reflects the response to macroeconomic news and rumors. Engle and Patton (2001), and Poon and Granger (2003) stress that volatility surface has empirically been proved to have persistence for a long time against market shocks. The long-memory phenomenon in time series is first considered by Hurst (1951, 1956). In these studies, he explains the long-term storage requirements of the Nile River. He shows that the cumulated water flows in a year depend not only on the water flows in recent years, but also on water flows in years much earlier prior to the present year. Mandelbrot and Van Ness (1968) explain and advance Hurst's studies by employing fractional Brownian motion. In analogy with Mandelbrot and Van Ness (1968), Granger (1980), Granger and Joyeux (1980) and Hosking (1981) develop Autoregressive Fractionally Integrated Moving Average (ARFIMA) models to account for the long memory in time series data. However, an empirical study regarding the usefulness of ARFIMA model is conducted by Bhardwaj and Swanson (2006), who find strong evidence in favor of ARFIMA in absolute, squared and log-squared stock index returns. In this regard, Ding and Granger (1996) point out that a number of other processes can also have the long-memory property. Further, a fractionally integrated generalized autoregressive conditional heteroskedasticity (FIGARCH) is primarily developed by Baillie, Bollerslev and Mikkelsen (1996), but later modified by Chung (1999). Nevertheless, Quoreshi (2014) develops an Integer-valued ARFIMA (INARFIMA) model to account for the long-memory property in a high frequency count data framework.

This paper incorporates conditional heteroscedasticity properties in the long memory model and applies the model on squared returns of BRICS (Brazil, Russia, India, China, and South Africa), UK and USA. The new model is called Fractionally Integrated Moving Average Conditional Heteroskedasticity (FIMACH). This model is designed, in a similar fashion to Quoreshi (2014), for non-integer data. The main difference between the introduced model and model in ARFIMA class is that this model class can study the heteroskedasticity property on the level series, while the ARFIMA-FIGARCH class studies the same on the fractionally differenced series through Fourier transformation. One obvious advantage of the FIMACH model over the ARFIMA-FIGARCH class is that the model can easily be extended to multivariate settings for the level series. The model

may additionally be used to measure the reaction times for macro-economic news or rumors, and captures information spread through the system. The model is specified in terms of first and second order moments conditioned on historical observations. We perform a Monte-Carlo simulation, where we find that ARFIMA or FIGARCH is not suitable for data that are generated according to the FIMACH model. Empirically, we find evidence of long memory for squared stock return of UK, USA and BRICS countries. It is also found that the FIMACH model outperforms both FIGARCH and ARFIMA models in terms of eliminating serial correlations.

The paper is organized as follows. The ARFIMA-FIGARCH model class is discussed, and the FIMACH model is introduced in section 2. The estimation procedure of FIMACH is discussed in section 3. Section 4 presents a brief Monte Carlo experiment. The description of the empirical data is presented in section 5. The empirical results on the stock return volatilities are presented in section 6, and the concluding comments are included in section 7.

2. Model

We assume that $r_t = p_t - p_{t-1}$ is a stock index return time series, where p_t is price for the index at time t. Let σ_t^2 be the degree of index return volatility, proxied by the squared return r_t^2, which has a slow decaying autocorrelation function. The moving average representation of ARFIMA (0,d,0) of the series y_t is

$$\sigma_t^2 = u_t + d_1 u_{t-1} + d_2 u_{t-2} + d_3 u_{t-3} \ldots$$

or

$$\sigma_t^2 = (1+L)^{-d} u_t. \qquad (1)$$

where $y_t = \sigma_t^2$, t = 1, . . . , T time intervals and σ_t^2 has long memory properties. Note that σ_t^2 has long memory in a sense that the variable has a slow decaying autocorrelation function and the parameters $d_i = \Gamma(i + d)/[\Gamma(i+1)\Gamma(d)], \quad i = 0, 1, 2, \ldots$ where $d_0 = 1$. The u_t is i.i.d. sequence of random variables with unconditional mean $E(u) = \lambda$ and variance $V(\alpha u)$=$\alpha^2 \emptyset^2$ where $V(u) = E(u)^2$-λ^2=$\emptyset^2$. Conditionally, it

holds that $E(u|u) = u$ and $V(\alpha u|u)=\alpha^2 V(u|u)$ where $V(u|u) = u^2 - 2\lambda u + \lambda^2$. The conditional mean and variance for the moving average representation of ARFIMA (0,d,0) are

$$E(\sigma_t^2 |Y_{t-1}) = E_{t-1} = \lambda + \sum_{i=1}^{\infty} d_i \, u_{t-i} \qquad (2a)$$

$$V(\sigma_t^2 |Y_{t-1}) = V_{t-1} = \emptyset^2 + \sum_{i=1}^{\infty} d_i^2 \left(u_{t-i}^2 - 2\lambda u_{t-i} + \lambda^2\right). \qquad (2b)$$

where,Y_{t-1} is the information set available at time *t-1*. The conditional mean and variance vary with u_{t-i}. Since the conditional variance varies with u_{t-i}, there is a conditional heteroskedasticity property of moving average type that Brännäs and Hall (2001) called MACH(q). As λ and $\emptyset^2$ are not functions of time and $|\sum_{i=1}^{\infty} d_i| \leq \left|\sum_{i=1}^{\infty} d_i^2\right|$ for $d \in [-1, 1]$, it is sufficient that $\sum_{i=1}^{\infty} d_i < \infty$ for $\{\sigma_t^2\}$ to be a stationary sequence. We call the model Fractionally Integrated Moving Average Conditional Heteroskedasticity FIMACH (*d*) where *d* represents the long memory parameter. The main difference between this model and the model in the ARFIMA class is that this model can study the heteroskedasticity property on the level series, while, e.g., FIGARCH of ARFIMA class studies the same on the fractionally differenced series through Fourier transformation. The autocorrelation functions of the ARFIMA model class are assumed to be a hyperbolic function, while the general mathematical expression of the autocorrelation function for FIMACH is considerably complicated to derive, although possible. Assuming $E(u_t u_t|Y_{t-1}) = u_t^2$ and $E(u_t u_{t-i}|Y_{t-1}) = 0$ where $i = 1, 2, \ldots, \infty$, we can provide a simple form of conditional auto-correlation function at lag *k* for FIMACH as

$$\rho_{k|t-1} = \frac{\sum_{i=0}^{\infty} d_i d_{k+i} u_{t-i-k}^2}{V(\sigma_t^2 |Y_{t-1})} \qquad (2c)$$

where $k = -j, j$ and represent lag, and $d_0 = 1$. Note that this autocorrelation function varies with u_{t-i} which captures the heteroscedasticity property in autocorrelation function. The heteroscedasticity in autocorrelation

function for absolute return of stock is illustrated by Ding et al. (1993), although the authors assume a smooth function for explaining the autocorrelation. The model can be extended with random parameters as

$$\sigma_t^2 = u_t + d_1 u_{t-1} + d_2 u_{t-2} + d_3 u_{t-3} \ldots + \sum_{i=1}^{p} \theta_i \, u_{t-i} \tag{3}$$

where d_i capture the long memory properties and have the same definitions as in equation 1. The θ_i , $i = 1,2, \ldots$, comprise the random parameters and are independent of each other. These parameters capture the short term deviation from the long memory trend. We name this model FIMACH(*d, p*) model. The conditional mean and variance of the random coefficients of FIMACH(*d, p*) representation can be written as

$$E(\sigma_t^2 \,|Y_{t-1}) = E_{t-1} = \lambda + \sum_{i=1}^{\infty} d_i \, u_{t-i} \quad + \sum_{i=1}^{p} \theta_i \, u_{t-i} \tag{4a}$$

$$V(\sigma_t^2 \,|Y_{t-1}) = V_{t-1} = \emptyset^2 + \sum_{i=1}^{\infty} d_i^2 \left(u_{t-i}^2 - 2\lambda u_{t-i} + \lambda^2\right) + \sum_{i=1}^{p} \theta_i \left(u_{t-i}^2 - 2\lambda u_{t-i} + \lambda^2\right). \tag{4b}$$

Note that the moments are conditioned only on the previous observations, Y_{t-1}. The same stationary condition is applicable as for equation (2). The model can be used to measure mean and median reaction time to macroeconomic news and rumors[3]. The model can easily be extended to a multivariate setting. Hence, the covariance and Granger-Causality between two or several series can easily be studied in the same fashion as the VARMA model. These possibilities are limited in the FIGARCH of ARFIMA class, at least on the level series.

[3]This is the reaction to macroeconomic news/rumours in the (u_{jt}) sequence, we use the mean lag $\sum_{i=0}^{q_j} i \, \alpha_{ji}/w$, where$w = \sum_{i=0}^{q_j} \alpha_{ji}$ and $\alpha_{j0} = 1$ (see Quoreshi, 2012).

3. Estimation

If we do not assume a full density function, we may estimate the Quasi Maximum Likelihood (QML) Estimator as discussed by Weiss (1986) and Bollerslev and Wooldridge (1992) instead of Maximum Likelihood (ML) Estimator. Conditional Least Square (CLS), Feasible Generalized Least Square (FGLS), Generalized Methods of Moments (GMM) and possibly others, e.g. Two Stage Least Square (2SLS), are candidates for estimation. In the previous studies, it turns out that FGLS is the best estimator among the three in terms of eliminating serial correlation (Quoreshi, 2014). The CLS comes in the second position, which is almost as good as FGLS. Here, we only consider CLS, FGLS and ML class for estimation.

The Conditional least square (CLS) estimator for FIMACH(*d, p*) representation model have the following residual

$$e_t = y_t - E_{t-1} = \sigma_t^2 \; - \lambda - \sum_{i=1}^{\infty} d_i \, u_{t-i} \; - \sum_{i=1}^{p} \theta_i \, u_{t-i} \tag{5}$$

and the criterion function $S_{CLS} = \sum_{i=m+1}^{T} e_t^2$is minimized with respect to unknown parameters, i.e. $\psi = (\lambda, \theta' \text{and } d')$ where θ' and d' are vector of parameters with elements θ_i respective d_i. Using a finite maximum lag *m* in (5) instead of infinite lags may cause biasing effects. Due to omitted variables, i.e. $u_{t-m-1}, \dots, u_{t-\infty}$, we may expect a positive bias on the parameters λ, θ_iandd_i (Brännäs and Quoreshi, 2010). These moment conditions correspond to the normal equations of the CLS estimator that focuses on the unknown parameters of the conditional mean function. Alternatively and equivalently, the properties E(e_t) = 0 and E($e_t e_{t-j}$) = 0, $j \geq 1$ could be used. Note that the moment conditions for FIMACH(*d, 0*) can be obtained by setting $\theta_i = 0$.

The FGLS estimator minimizes

$$S_{FGLS} = \sum_{t=m+1}^{T} e_t \, \hat{V}^{-1} \tag{6}$$

with $\hat{V}^{-1}$ as given. The variance of error from CLS estimates may be used for approximation of $\hat{V}^{-1}$ in equation (6). Alternatively, $\hat{V}^{-1}$ can be estimated as specified in (4b) by employing estimates from CLS. The covariance matrix estimators for CLS and FGLS are

$$Cov(\hat{\psi}_{CLS}) = \left(\sum_{t=m+1}^{T} \frac{\partial e_t}{\partial \psi} \frac{\partial e_t}{\partial \psi'} \right)^{-1}$$

$$Cov(\hat{\psi}_{FGLS}) = \left(\sum_{t=m+1}^{T} \hat{V}^{-1} \frac{\partial e_t}{\partial \psi} \frac{\partial e_t}{\partial \psi'} \right)^{-1}.$$

The ML or QML estimator for FIMACH(*d, p*) representation model have the same residual as in equation (5), and maximize the following criterion function

$$f(\sigma_1^2, \sigma_2^2 \ \ldots, \sigma_T^2 \,|Y_{t-1}, \lambda, \theta_i \text{and } d_i) = \prod_{t=1}^{T} f(\sigma_t^2 \,|Y_{t-1}, \psi_i) \ = \left(\frac{1}{2\pi V_{t-1}}\right)^{T/2} exp\left(-\frac{\sum_{t=m+1}^{T} e_t^2}{2V_{t-1}}\right) (6)$$

Where $\psi_i = (\lambda, \theta_i \text{ and } d_i)$ and V_{t-1} is as in equation$(4b)$. Taking logarithm of equation (6), we may simply use the criterion function and minimize the function as

$$Lnf\left(\sigma_1^2, \sigma_2^2 \ \ldots, \sigma_T^2 \,|Y_{t-1}, \lambda, \theta_i, d_i \text{and} \hat{V}_{t-1}\right) = -\frac{T}{2}\ln\left(\hat{V}_{t-1}\right) - \ln(2\pi) - \left(\frac{\sum_{t=m+1}^{T} e_t^2}{2\hat{V}_{t-1}}\right) \quad (7)$$

where $\hat{V}_{t-1}$is an estimate for V_{t-1} that is to be estimated. Since T, 2 and π are constants, we can equivalently minimize the following criterion function

$$Lnf\left(\sigma_1^2, \sigma_2^2 \ \ldots, \sigma_T^2 \,|Y_{t-1}, \lambda, \theta_i, d_i \text{and} \hat{V}_{t-1}\right) = -\ln(\hat{V}_{t-1}) - \left(\frac{\sum_{t=m+1}^{T} e_t^2}{\hat{V}_{t-1}}\right). \quad (8)$$

Note that the $\hat{V}_{t-1}$ is to be estimated at the same time as the other parameters. If the estimation is sensitive to the start value of $\hat{V}_{t-1}$, we can obviously estimate CLS at the first stage and calculate the $\hat{V}_{t-1}$ which can be used as the start value for QML. We call this estimation procedure Two Stage Quasi Maximum Likelihood (2SQML) Estimation. The covariance matrix estimators for QML and 2SQML are

$$Cov(\hat{\psi}_{QML}) = \left(\sum_{t=m+1}^{T} \hat{V}^{-1} \frac{\partial e_t}{\partial \psi} \frac{\partial e_t}{\partial \psi'} \right)^{-1}$$

$$Cov(\hat{\psi}_{2SQML}) = \left(\sum_{t=m+1}^{T} \hat{V}^{-1} \frac{\partial e_t}{\partial \psi} \frac{\partial e_t}{\partial \psi'} \right)^{-1}.$$

4. Monte Carlo experiment

Smith et al. (1996) and Quoreshi (2014) have studied the bias and misspecification in ARFIMA respective INARFIMA models. Drost et al. (2009) investigated finite sample behavior of semiparametric integer-valued AR(p) models, while Brännäs and Quoreshi (2010) studied finite lag misspecification when the data is generated according to an infinite-lag INMA model. In this brief Monte Carlo experiment we study the bias, MSE, Ljung–Box statistics, AIC and SBIC properties of the ML estimators for finite-lag specifications, when data is generated according to FIMACH (*d,0*). The data generating process is as in (1), with $d = 0.1, 0.25 \text{ and } 0.4$ and lag length $m = 70$. The u_t sequence is generated from i.i.d. normal distribution, with mean 25 and standard deviation 10. Six time series with length $T = 4500 \text{ and } T = 900$ are generated. The first 500 observations are discarded to avoid the start-up effect. The results for the Monte Carlo experiment are given in Table 1. We also generate four other series in a similar fashion, with mean 25 and standard deviation 4 and 100 to study the performance of CLS, FGLS, 2SQML and QML estimators. We also evaluate the FIMACH(*d, 0*), ARFIMA (*0, d, 0*), FIGARCH (1,1) and GARCH (1,1) models in terms of eliminating serial correlations when the data are generated in accordance with FIMACH(*d, 0*). These results are presented in Table 2.

We set $\hat{\lambda}$ equal to mean values for the generated series instead of $\lambda = 25$. By doing so, we eliminate biased effect of λ in the generated series. Hence, we can study the bias of the estimated parameter d due to misspecification of lag length m more appropriately. The Monte Carlo study shows that as m increases towards $m = 70$, the bias in d decreases (see Table 1). For m less than 70 we find positive bias in d, while the bias is negative for $m = 90$. Biases are smaller for $T = 4500$ when m is less than 70 than those for $T = 9000$. MSE decreases as sample size increases or m increases. Like Brännäs and Quoreshi (2010) and Quoreshi (2014), we conclude that we may expect a positive biasing effect on the parameters due to omitting variables, i.e. $u_{t-m-1}, \dots, u_{t-\infty}$. The statistics for AIC and SBIC decrease as lag length increases, and are noted lowest at $m = 90$. Hence, the standard AIC and SBIC need to be corrected in order to choose optimal lag lengths. As expected, the LB statistics are, with some exceptions, lowest at $m = 70$. The exceptions may arise due to the conditional heteroskedasticity nature of the data set. Hence, it is appropriate to evaluate at more than one time point or to use an average value of LB statistics for a number of time points.

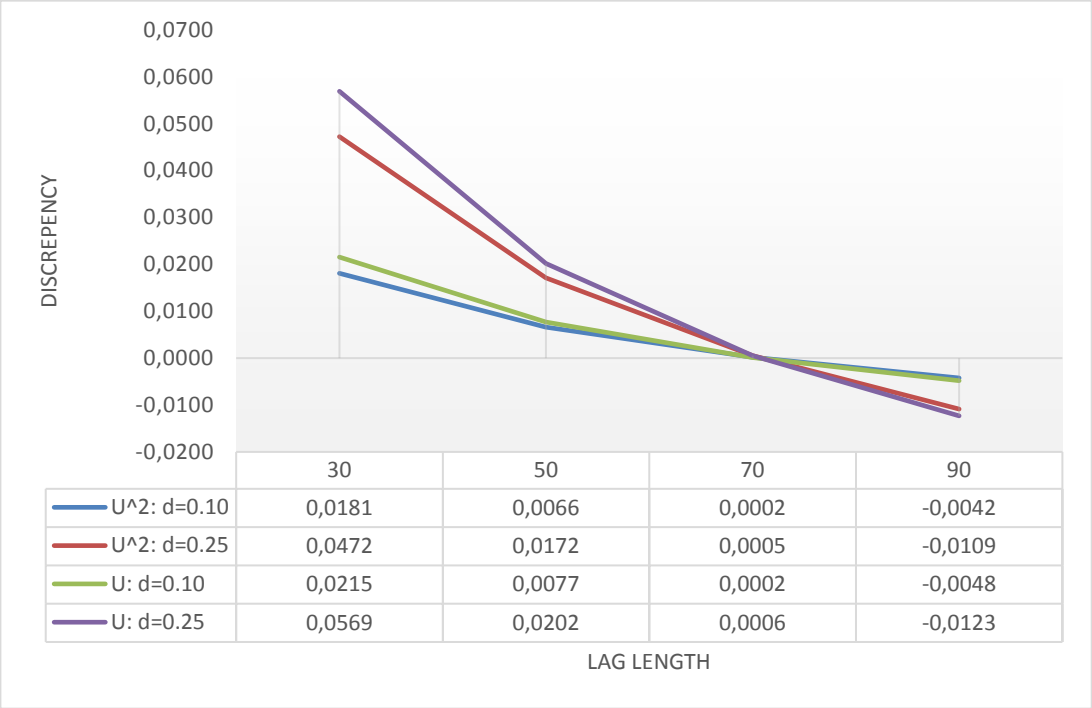
Insert Table 1 about here

When the data is generated according to FIMACH $(d, 0)$, it is not appropriate to use ARFIMA, FIGARCH or GARCH model (see Table 2). The ARFIMA model reduces the serial correlation successfully, but it does not perform as well as FIMACH. This shows that there is need of using FIMACH instead of ARFIMA when we need to take account of the heteroskedasticity property in the long memory. FIGARCH and GARCH take account of heteroskedasticity in the short memory. Hence, these models did not perform well. FGLS and CLS perform consistently well, and somewhat better than QML. It turns out that QML is sensitive to start values. Estimating CLS at the first stage and using the CLS estimates as start value for QML estimator, we estimate 2SQML which performs best out of these estimators.

Insert Table 2 about here

We have also conducted similar Monte-Carlo studies with innovation U^2. The results are quite similar. Average discrepancy (d̂-d) with number of simulation equals 100 for FIMACH time series, with innovation U^2 respective U; when the data are generated with lag length 70 and d=0.1 respective d=0.25 are presented in Figure 1 below.

Figure 1: Average discrepancy ($\widehat{\boldsymbol{d}} - \boldsymbol{d}$) with number of simulation equals 100 for FIMACH time-series with innovation U^2 respective U when the data are generated with lag length 70 and $\boldsymbol{d} = \mathbf{0.1}$ respective $\boldsymbol{d} = \mathbf{0.25}$.



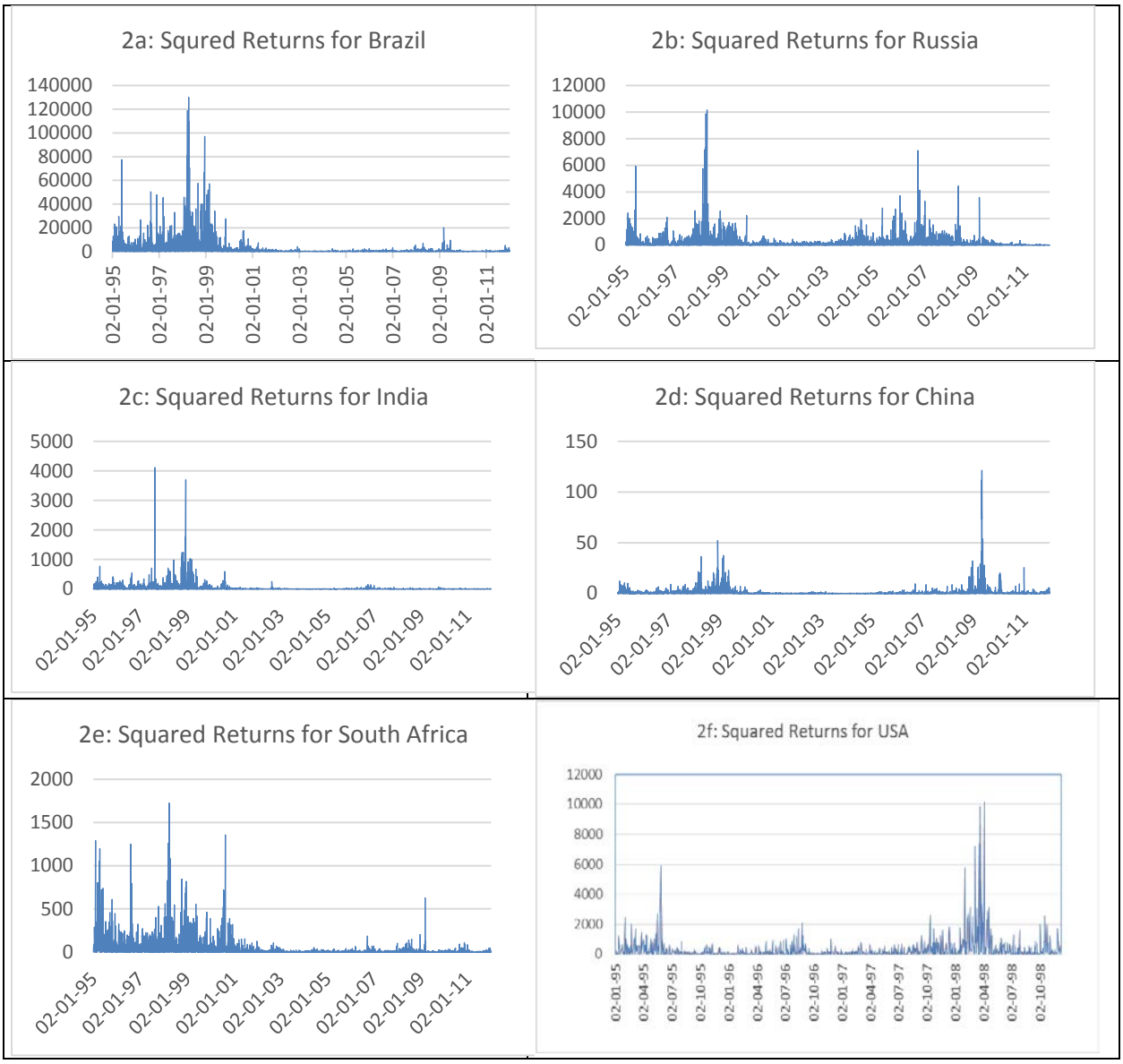
5. Empirical Data

The daily squared stock MSCI index return series over the period of 17 years from 1995 to 2011 for the UK, USA and BRICS are applied in this paper (see Figure 2 below). Each series comprises 4435 observations. The descriptive statistics of the data set are given in Table 3. The mean squared stock index return for Brazil is about 1800, which is the largest among the BRICS counties. The corresponding number for China is about 1, which is the smallest among the BRICS countries. The corresponding mean for USA and UK are 172 and 195

respectively. The skewness, kurtosis and Jarque-Bera statistics indicate that the data are not from normal distribution. The autocorrelation functions are presented in Figure 3 below. The autocorrelation functions for all of the series decay very slowly, which indicates a long memory behavior in the squared return series.

Insert Table 3 about here

Figure 2: The daily squared stock index return series over the period of 17 years from 1995 to 2011 for the BRICS countries, USA and UK.



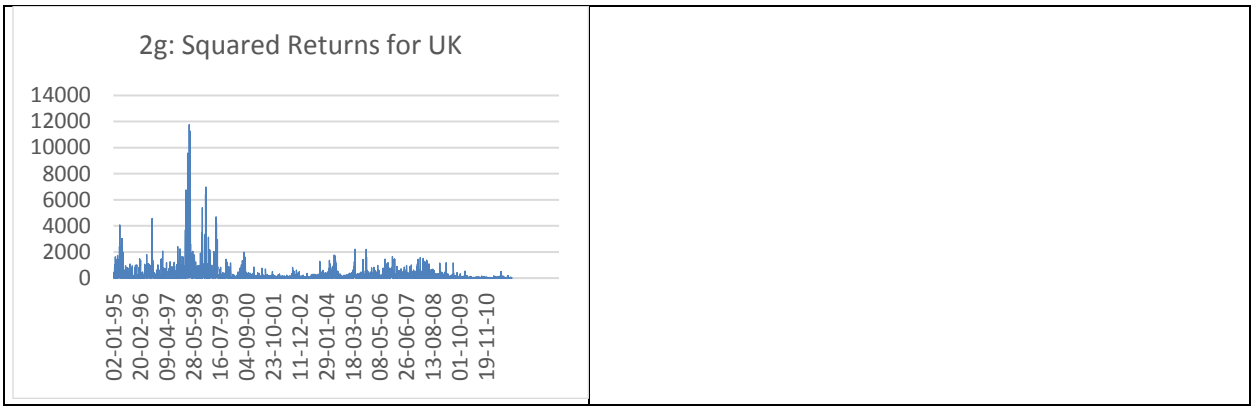
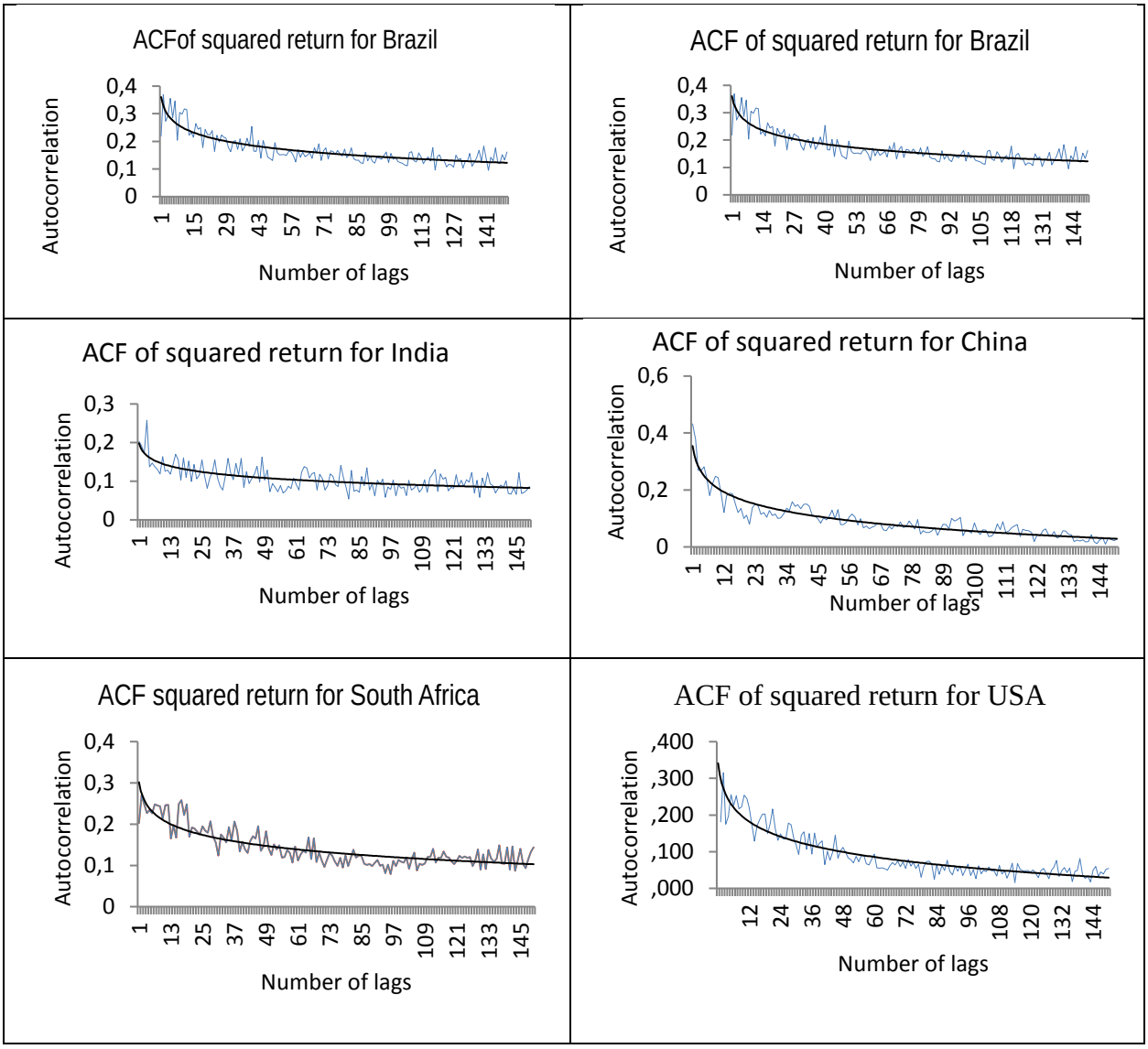
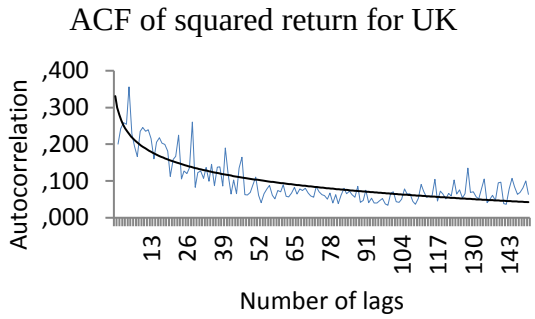


Figure 3: Autocorrelation Function for BRICS countries, USA and UK.





6. Empirical Results

The empirical results of the squared stock index returns for USA, UK and BRICS countries are presented in Table 4a-g. CLS, FGLS and QML estimators have been employed to evaluate the performance of the estimators. FIMACH (d,0), FIGARCH (0,d,0), FIGARCH (1,d,1) and GARCH (1,1) are estimated to find out the most suitable model for the volatility return. It turns out that CLS, FGLS and QML estimators have performed equally well for the time series of USA, Brazil, Russia and China in terms of reducing serial correlation. However, the QML estimator outperforms CLS and FGLS for the time series of India and South Africa, while CLS and FGLS perform better than QML estimator for the time series of UK. Hence, we conclude that QML is somewhat better estimator among these. It is to be noted that the performance the QML estimator is highly sensitive to the start values, and may turn out worse than that of FGLS or CLS if the start values are not selected carefully. We suggest that the value for the autocorrelation at lag one may be chosen as the start values for the fractional integration parameter. The variance of residuals from the CLS estimator may be used as the start value for $\hat{V}_{t-1}$ of the QML estimator which we call 2SQML. We also recommend to use both QML and CLS estimators in order to determine the best estimates in terms of reducing serial correlation.

We find that the GARCH is not an appropriate model for the time series, as the estimated parameters deviate substantially from the expected values. The FIMACH (0,*d*) turns out to be the best in terms of eliminating serial correlations. This model performs much better than that of FIGARCH (0,d,0) and FIGARCH (1,d,1) for all

the five time series. For Brazil, the Ljung-Box (LB) statistics for residuals for FIMACH is about 1430, while the corresponding number for FIGARCH is about 16,882 (see Table 4a). The corresponding statistics for USA are 698 for FIMACH and 7147 for FIGARCH (see Table 4f). The Ljung-Box statistics for standardized residuals for FIGARCH are smaller than that of residuals, but these are much larger than the corresponding statistics for FIMACH. Note that the Ljung-Box statistics for standardized residuals, and residuals are the same for all FIMACH estimations. This may imply that the FIMACH model captures heteroskedasticity properly.

Employing the proposed FIMACH $(d, 0)$ model, we find that the squared stock return index for each of the USA, UK and BRICS countries have long memory properties. The persistence in terms of days varies among the countries. The effects of macroeconomics news and rumors on stock market volatility for India persist up to 35 days, while the corresponding number of days for South Africa is 70. The stock market in China reacts initially more than any other BRICS countries, since the fractional integration parameter (*d*) for China is 0.283, which is the largest among the BRICS countries. The stock market in Russia reacts initially least ($d = 0.167$) among those countries. The volatility intensity increases for all the seven stock markets when the macro-economic news or rumors break out, and the impact remains between 35 and 70 days, and fades away very slowly with time.

Insert Tables 4a-g about here

7. Conclusion

This paper introduces a new class of long memory model for volatility of stock returns. The model introduced is capable of taking account of heteroskedasticity in long memory. The conditional first- and second-order moments are provided. The CLS, FGLS and QML are discussed and 2SQML estimator is proposed. In Monte Carlo experiments we find that it is not appropriate to use ARFIMA, FIGARCH or GARCH model if the data is generated according to FIMACH $(d, 0)$. The ARFIMA model reduces the serial correlation successfully, but it does not perform as well as FIMACH. From the empirical results, we establish that the squared returns of stock index for the BRICS countries, UK and USA have long memory properties. However, the effects of macroeconomics news and rumors on stock return volatility vary among the countries. We also find that the volatility intensity increases for all the seven stock markets when the macro-economic news or rumors break

out, and that the impact remains between 35 and 70 days and fades away very slowly with time. CLS and FGLS estimators perform equally well in terms of residual properties, while the QML estimator performs in somewhat better among the three estimators. The results of the simulation study indicate that 2SQML performs best among the five estimators and hence 2SQML is suggested to be used when QML does not perform well. Both in simulation and empirical studies, we find that the proposed model FIMACH outperforms FIGARCH in terms of eliminating serial correlations.

Acknowledgement:

We acknowledge financial support from NASDAQ OMX Nordic Foundation.

References:

Baillie, R. T., Bollerslev, T. and Mikkelsen, H.O., 1996.Fractionally integrated generalized autoregressive conditional hetroskedasticity. Journal of Econometrics 74, 3-30.

Bhardwaj, G., and Swanson, N.R., 2006. An Empirical Investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. Journal of Econometrics 131(1&2), 539-578.

Bollerslev, T. and J.M. Wooldridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, Econometric Reviews 11, 143-172.

Brännäs, K. and Hall, A., 2001. Estimation in integer-value moving average models. Applied Stochastic Models in Business and Industry 17, 277-291.

Brännäs, K. and Quoreshi, A.M.M.S., 2010. Integer-value moving average modeling of the transactions in stocks. Applied Financial Economics 20, 1429-1440.

Chung, C. F., 1999. Estimating the Fractionally Integrated GARCH Model. National Taiwan University, Working Paper.

Ding, Z., Granger, C.W.J. and Robert F. Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83-106.

Engle, R.F & Patton, A.J., 2001. What good is a Volatility Model? Quantitative Finance 1, 237-245

Poon, S.H. and Granger, C.W.J., 2003. Forecasting volatility in financial markets: A review.Journal of Economic Literature 41, 478–539.

Granger, C.W.J., 1980. Long Memory Relationships and the Aggregation of Dynamic Models.Journal of Econometrics 14, 227-238.

Granger, C.W.G., Joyeux, R., 1980.An introduction to long memory time series models and fractional differencing.Journal of Time Series Analysis 1, 15-29.

Hosking, J., 198l. Fractional differencing, Biometrika 68 (l), 165-176.

Quoreshi, A.M.M.S., 2014. Long-memory integer-valued time series model. Quantitative Finance Vol 12, 2225-2235.

Weiss, A.A., 1986. Asymptotic theory for ARCH models: Estimation and testing, Econometric Theory 2, 107-131.

Table 1: Bias, MSE, Ljung–Box statistics, AIC and SBIC properties of the ML estimators for finite-lag specifications, when data is generated according to FIMACH (0,d,0) Model with $d = 0.1, 0.25$ and 0.4 and $m = 70$ and σ^2=100.

		T=4500			T=9000		
Lag	Parameters	**0.10**	**0.25**	**0.40**	**0.10**	**0.25**	**0.40**
	δ	0.168***	0.432***	0.698***	0.167***	0.431***	0.696***
	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MSE	99.115	104.898	130.848	97.579	103.611	129.173
	LB100	130.669	297.170	1047.228	166.566	525.401	2045.455
	LB200	236.154	417.668	1214.240	285.267	651.793	2200.088
	AIC	20659.704	20907.183	21811.986	41201.319	41723.602	43516.296
M10	SBIC	20730.210	20977.689	21882.492	41279.461	41801.744	43594.438
	δ	0.122***	0.308***	0.496***	0.122***	0.307***	0.495***
	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MSE	98.180	99.305	103.427	96.901	97.936	101.736
	LB100	101.855	116.564	167.503	108.976	151.998	284.131
	LB200	202.485	219.121	268.172	225.197	268.950	405.750
	AIC	20566.007	20616.690	20797.669	41048.905	41143.489	41482.381
M30	SBIC	20596.463	20815.250	20996.228	41127.023	41221.607	41560.499
	δ	0.108***	0.271***	0.434***	0.108***	0.270***	0.433***
	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
M50	MSE	98.177	100.234	104.320	97.025	97.903	100.766

	LB100	101.107	111.853	150.678	105.672	116.088	180.584
	LB200	200.968	217.921	264.321	224.083	234.140	299.255
	AIC	20514.152	20606.353	20784.158	41049.001	41129.531	41387.421
	SBIC	20840.585	20932.787	21110.592	41411.071	41491.601	41749.491
	δ	0.101***	0.251***	0.402***	0.100***	0.250***	0.401***
	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.001)
	MSE	98.027	99.939	105.527	96.821	97.604	100.086
	LB100	103.897	115.392	160.502	103.300	108.986	134.357
	LB200	200.740	212.267	264.265	222.201	226.818	249.751
	AIC	20455.626	20541.172	20782.114	40989.067	41060.157	41281.722
M70	SBIC	20909.753	20995.299	21236.241	41492.967	41564.057	41785.621
	δ	0.096***	0.238***	0.380***	0.095***	0.237***	0.379***
	(s.e.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MSE	98.110	100.157	106.060	96.917	97.794	100.522
	LB100	110.666	162.701	332.128	108.179	143.667	292.239
	LB200	210.420	267.350	445.421	228.784	266.039	422.550
	AIC	20407.668	20498.686	20750.943	40936.006	41016.225	41261.143
M90	SBIC	20989.306	21080.324	21332.581	41581.644	41661.864	41906.782

Table 2: Comparing between FIMACH, FIGARCH and GARCH and between QMLE, 2SQMLE, FGLS and CLS estimators, when data is generated according to FIMACH $(\mathbf{0}, \boldsymbol{d}, \mathbf{0})$ models with $\boldsymbol{d} = \mathbf{0.25}$ and $\boldsymbol{m} = \mathbf{70}$.

		T=4500				T=9000			
Lag	Ljung-Box	FIMACH	ARFIMA	FIGARCH	GARCH	FIMACH	ARFIMA	FIGARCH	GARCH
σ=4	LB_{100}	90.521	138.064	2704.712	2730.476	90.521	138.064	1455.198	1578.196
	LB_{200}	198.883	247.337	2781.968	2805.809	198.883	247.337	1564.312	1687.964
σ=10	LB_{100}	90.521	138.064	2704.712	2730.476	90.521	138.064	1455.198	1578.196
	LB_{200}	198.883	247.337	2781.968	2805.809	198.883	247.337	1564.312	1687.964
σ=100	LB_{100}	90.521	138.064	1455.198	1578.196	90.521	138.064	1455.198	1578.196
	LB_{200}	198.883	247.337	1564.312	1687.964	198.883	247.337	1564.312	1687.964
		QMLE	2SQMLE	FGLS	CLS	QMLE	2SQMLE	FGLS	CLS
σ=4	LB_{100}	679.924	90.521	100.409	100.409	291.590	90.521	291.586	291.592
	LB_{200}	843.340	198.883	200.232	200.232	406.221	198.883	406.217	406.224
σ=10	LB_{100}	109.710	90.521	100.229	100.229	108.557	90.521	108.557	108.557
	LB_{200}	206.859	198.883	199.617	199.617	226.373	198.883	226.373	226.373
σ=100	LB_{100}	99.964	90.521	100.139	100.139	102.549	90.521	102.549	102.549
	LB_{200}	198.430	198.883	199.260	199.260	221.728	198.883	221.728	221.728
		σ=4	σ=10	σ=100		σ=4	σ=10	σ=100	
Generated series	LB_{100}	1885.1	1885.1	1885.1		1885.1	1885.1	1885.1	
	LB_{200}	1991.9	1991.9	1991.9		1991.9	1991.9	1991.9	

Table 3: Descriptive Statistics

This table reports descriptive statistics of squared returns for BRICS (Brazil, Russia, India, China, and South Africa) countries, UK and US for the period 1995-2011.

Country	Mean	Median	Std. Dev.	Minimum	Maximum	Skewness	Kurtosis	Jarque-Bera	Observations
Brazil	1799.718	158.722	6476.492	0.000	130113.100	9.301	125.839	2851677.000***	4435
Russia	216.787	23.981	949.744	0.000	35765.620	19.805	594.868	65009317.000***	4435
India	26.928	2.151	119.300	0.000	4115.351	18.851	542.786	54093083.000***	4435
China	1.032	0.123	4.089	0.000	121.396	14.887	341.988	21393884.000***	4435
South Africa	33.414	4.351	95.426	0.000	1725.239	7.215	78.675	1096490.000***	4435
UK	195.152	46.909	548.819	0.000	11764.870	10.107	152.599	4210143.000***	4435
US	171.966	34.018	472.666	0.000	10158.220	9.380	141.160	3591553.000***	4435

Table 4a: Empirical results for squared stock return index for Brazil.

Coefficients	FIMACH (*d,0*)			FIGARCH (*0,d,0*)	FIGARCH (1,d,1)	GARCH (1,1)
	QMLE Estimates (s.e.)	FGLS Estimates (s.e.)	CLS Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)
Constant (Mean)	737.541***	737.541***	737.541***	129.996***	126.228***	101.091***
	(72.54)	(73.65)	(72.83)	(8.71)	(6.28)	(8.41)
Constant (Variance)					-46632.882***	634.873
					(1915.54)	(487.658)
d	0.208***	0.208***	0.208***	0.661***	0.751***	
	(0.029)	(0.029)	(0.028)	(0.004)	(0.003)	
α						1.081***
						(0.03)
β					0.513***	0.879***
					(0.00)	(0.02)
VAR	34453328***	34468875.171	34468875.171	46284279	41936202	41927397
	(5577746)					
Q(100)	1430.501***	1430.501***	1430.501***	16882.451***	16882.451***	16881.811***
Q(200)	2345.913***	2345.913***	2345.913***	25714.538***	25714.538***	25713.570***
LB_{SD} (100)	1430.501***	1430.501***	1430.501***	5142.817***	4164.267***	1712.273***
LB_{SD} (200)	2345.913***	2345.913***	2345.913***	8786.100***	6739.816***	2668.345***
Lag length	45	45	45			

Table 4b: Empirical results for squared stock return index for Russia.

Coefficients	FIMACH(d,0)			FIGARCH (*0,d,0*)	FIGARCH(1,d,1)	GARCH (1,1)
	QMLE Estimates (s.e.)	FGLS Estimates (s.e.)	CLS Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)
Constant (Mean)	107.259*** (12.47)	107.259*** (14.48)	107.258*** (14.22)	24.075*** (1.86)	10.409*** (0.276)	2.901 (1.84)
Constant (Variance)					-1860.043*** (61.58)	5.340 (7.66)
d	0.167*** (0.04)	0.167*** (0.04)	0.167*** (0.04)	0.615*** (0.00)	0.728*** (0.00)	
α						1.302*** (0.10)
β					0.537*** (0.00)	0.808*** (0.03)
VAR	823353*** (309982.45)	823727	823727		901821	901627
LB (100)	997.362***	997.362***	997.362***	4487.831***	4487.831***	4488.051***
LB (200)	1374.943***	1374.943***	1374.943***	5959.304***	5959.304***	5959.476***
LB_{SD} (100)	997.362***	997.362***	997.362***	3363.808***	1348.205***	112.467***
LB_{SD} (200)	1374.943***	1374.943***	1374.943***	5564.297***	2043.733***	232.909***
Lag length	41	41	41			

Table 4c: Empirical results for squared stock return index for India

	FIMACH			FIGARCH (*0,d,0*)	FIGARCH (1,d,1)	GARCH (1,1)
	QML Estimates (s.e.)	FGLS Estimates (s.e.)	CLS Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)
Constant (Mean)	12.709*** (2.57)	17.614*** (1.59)	17.614*** (1.59)	1.499*** (0.11)	1.579*** (0.11)	1.469*** (0.22)
Constant (Variance)					0.311* (0.18)	0.035 (0.23)
d	0.185*** (0.05)	-0.169*** (0.04)	-0.169*** (0.04)	0.549*** (0.11)	0.999*** (0.00)	
α			0.319 (0.18)			1.204*** (0.10)
β					0.934*** (0.02)	0.854*** (0.04)
VAR	12829.171*** (4439.03)	13166.002	13166.002	14229.523	14229.523	14226.460
LB (100)	717.844***	1572.995***	1572.995***	6214.240***	6214.240***	6214.247***
LB (200)	1150.830***	2582.103***	2582.103***	9987.077***	9987.077***	9987.540
LB_{SD} (100)	717.844***			1319.669***	1052.742***	1057.400***
LB_{SD} (200)	1150.830***			2479.399***	1840.481***	1791.520***
Lag length	35	35	35			

Table 4d: Empirical results for squared stock return index for China.

Coefficients	FIMACH			FIGARCH (0,d,0)	FIGARCH (1,d,1)	GARCH (1,1)
	QML Estimates (s.e.)	FGLS Estimates (s.e.)	CLS Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)
Constant (Mean)	0.283***	0.283***	0.283***	0.220***	0.077***	0.055***
	(0.11)	(0.10)	(0.10)	(0.02)	(0.01)	(0.01)
Constant (Variance)					-0.021***	-0.000
					(0.00)	(0.00)
d	0.312***	0.312***	0.312***	0.490***	0.714***	
	(0.11)	(0.11)	(0.11)	(0.00)	(0.00)	
α						1.078***
						(0.04)
β					0.684	(0.916)
					(0.00)	(0.02)
VAR	12.567***	12.569	12.569	16.716	16.716	16.713
	(3.08)					
LB (100)	437.879***	437.879***	437.879***	8062.879***	8062.879***	8063.962***
LB (200)	638.522***	638.522***	638.522***	9148.716***	9148.716***	9150.167***
LB_{SD} (100)	437.879***	437.879***	437.879***	4259.401***	1871.683***	820.221***
LB_{SD} (200)	638.522***	638.522***	638.522***	7478.183***	3316.763***	1354.586***
Lag length	42	42	42			

Table 4e: Empirical results for squared stock return index for South Africa.

	FIMACH			FIGARCH (0,d,0)	FIGARCH (1,d,1)	GARCH (1,1)
	QML Estimates (s.e.)	FGLS Estimates (s.e.)	CLS Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)
Constant (Mean)	12.782***	20.314***	20.314***	4.181***	3.274***	3.600***
	(1.37)	(1.28)	(1.27)	(0.298)	(0.15)	(0.28)
Constant (Variance)					-22.094***	0.587
					(1.12)	(0.43)
d	0.194***	-0.166***	-0.166***	0.533	0.689***	
	(0.02)	(0.02)	(0.02)	(0.00)	(0.00)	
α						1.115***
						(0.05)
β					0.572***	0.867***
					(0.00)	(0.03)
VAR	7036.353***	7381.231	7381.231	9104.279	9104.279	9102.426
	(996.96)					
LB (100)	772.396***	2684.763***	2684.764***	11702.479***	11702.479***	11693.869***
LB (200)	1266.134***	4513.014***	4513.015***	18138.175***	18138.175***	18126.195***
LB_{SD} (100)	772.396***	2684.763***	2684.764***	3285.210***	1606.489 ***	1393.617***
LB_{SD} (200)	1266.134***	4513.014***	4513.015***	5571.753***	2285.839***	1831.270***
Lag length	70	70	70			

Table 4f: Empirical results for squared stock return index for USA.

	FIMACH			FIGARCH (0,d,0)	FIGARCH (1,d,1)	GARCH (1,1)
	QML Estimates (s.e.)	FGLS Estimates (s.e.)	CLS Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)
Constant (Mean)	70.071***	70.071***	70.071***	50.926***	18.044***	13.701***
	(8.539)	(8.412)	(8.403)	(2.918)	(1.616)	(3.349)
Constant (Variance)					-1106.968***	-0.662
					(47.472)	(2.993)
d	0.187***	0.187***	0.187***	0.472	0.669***	
	(0.034)	(0.035)	(0.035)	(0.016)	(0.004)	
α						1.045***
						(0.018)
β					0.622***	0.915***
					(0.005)	(0.015)
VAR	192930.645***	192974.940	192974.940	223369.828	223369.828	223325.106
	(32020.920)					
LB (100)	697.590***	697.590***	697.590***	7147.323***	7147.323***	7145.920***
LB (200)	868.459***	868.459***	868.459***	8028.339***	8028.339***	8026.328***
LB_{SD} (100)	693.986***	697.590***	693.986***	4432.055***	2445.256***	1061.702***
LB_{SD} (200)	864.174***	868.459***	864.174***	6799.300***	3694.440***	1744.355***
Lag length	70	70	70			

Table 4g: Empirical results for squared stock return index for UK.

	FIMACH			FIGARCH (0,d,0)	FIGARCH (1,d,1)	GARCH (1,1)
	QML Estimates (s.e.)	FGLS Estimates (s.e.)	CLS Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)	QML Estimates (s.e.)
Constant (Mean)	123.736***	78.056***	78.056***	62.144***	36.174***	35.517***
	(7.816)	(10.074)	(10.176)	(4.688)	(1.988)	(0.360)
Constant (Variance)					-655.158***	45.497***
					(77.010)	(0.004)
d	-0.162***	0.192***	0.192***	0.444	0.682***	
	(0.027)	(0.035)	(0.035)	(0.018)	(0.007)	
α						1.039***
						(0.014)
β					0.660***	0.918***
					(0.007)	(0.013)
VAR	270893.898***	258868.144	258868.144	301142.683	301142.683	301080.523
	(52010.797)					
LB (100)	2086.949***	1012.821***	1012.821***	7671.396***	7671.396***	7670.816***
LB (200)	2984.945***	1499.144***	1499.144***	10420.796***	10420.796***	10420.151***
LB_{SD} (100)	2086.902***	1012.821***	1012.317***	3809.926***	2201.357***	1600.740***
LB_{SD} (200)	2980.607***	1499.144***	1498.329***	6070.656***	3661.447***	2832.232***
Lag length	70	70	70			

Probabilistic forecasting and simulation of electricity prices

Peru Muniain[1] * and Florian Ziel[2]

[1] University of the Basque Country, UPV/EHU. E-mail: peru.muniain@ehu.eus
[2] University of Duisburg-Essen. E-mail: florian.ziel@uni-due.de.

Abstract. In this paper we include dependency structures to electricity price forecasting and prediction evaluation. We work with off-peak and peak time series from the German-Austrian day-ahead price, hence we analyse a bivariate data. We first estimate the mean of both time series, and then in a second step we estimate the residuals. The mean equation is estimated by OLS and elastic net and the residuals are estimated by maximum likelihood. Our contribution is to include bivariate jump component on a mean reverting jump diffusion model in the residuals. The evaluation of the models' forecasts are made using four different criteria, among them we include the Energy score in order to measure whether the correlation structure between the time series is properly included or not. In the results it is observed that the models with bivariate jumps provide better results with the Energy score, which means that it is important to consider this structure to properly forecast correlated time series.

Keywords: Probabilistic Forecasting; Electricity prices; Jump diffusion model; Multivariate GARCH; Forecasting evaluation; Energy score

1 Introduction

In the last decades after the deregulation of the electricity markets it has become increasingly important to capture the uncommon features of electricity prices, such as nonstorability which makes electricity prices really volatile, see [1]. In this paper we use different time series models to forecast electricity by simulation and then we evaluate those forecasts with various criteria each one with their own properties. We believe that to forecast multivariate time series it is crucial to take into account the dependency structures in order to properly forecast the time series. The innovation in this paper is that we include dependency structures in some of the multivariate forecasting models as well as in one of the forecast evaluation criterion to show that the incorporation of the dependency structures substantially improve the electricity price forecasts.

As mentioned above, electricity prices exhibit special characteristics which are usually classified in literature, see [1] and [2]. These properties are especially i) mean reverting behaviour, ii) seasonal behaviour, iii) time depended volatility, iv) occurrence of price spikes and v) cross-period effects (e.g. night hours influence day-time hours even though they are realised at the same day). All the aspects are known in the literature. Still, there is no electricity price forecasting model which incorporates all these aspects. For instance, [3] cover all effects but no interaction effects, [4] consider all effects but price spike effects. We propose electricity price models which incorporate all the mentioned effects into a probabilistic electricity price forecasting framework.

In order to forecast the prices a two step approach is followed. In the first step the conditional mean model is estimated, the mean must be properly estimated so that the residuals have zero mean. Therefore, in the mean equation all the seasonal properties must be included, in this sense [5] and [6] propose mean equations with autoregressive, non-linear effects and seasonal effects. Once properly estimated the conditional mean model we proceed with the estimation of the residuals, the residuals must have zero mean so the models differ in the structures of the standard deviation. We consider mean reverting jump diffusion models (MRJD) such as the model included by [7] and applied to electricity prices. The MRJD model is a OrnsteinUhlenbeck (OU) process proposed by [8]. Unlike [9] and [10] where they first estimate the jump component and then they assume a OU process in the continuous part, we first estimate the mean model and then we apply assume MRJD structure in the residuals. [1] offer a great review applying MRJD models to the electricity price forecasting field. As mentioned above our interest is the dependency structure among different time series and we include this dependency structure assuming correlated jump occurrence processes, as far as we concern never assumed before. In order to have a correlated jump we focus on the bivariate Bernoulli process, proposed by [11].

* corresponding author

Once estimated the models electricity prices are simulated and forecasted, then these forecasts and its paths are evaluated using different criteria. In this article we use four different criteria: Mean absolute error (MAE), mean square error (MSE), pinball score (PB, also known as quantile loss) and energy score (ES). The first two evaluation criteria are the most used ones in the literature of forecasting evaluation, for instance, [7] apply MSE to evaluate the electricity price forecasts from a model including spikes and other kind of structures as ARIMA or GARCH. Furthermore, [12] use the MSE and MAE criteria to evaluate the performance of different electricity price forecasts in the NORDPOOL market. In this paper we focus more on the PB and ES as we are interested in the performance capturing the whole distribution and how the different models capture dependency structures. The PB has been applied by [13], [14] and [15], all of them participated in an electricity price forecasting competition and the PB was used to check performance as the objective was to approximate the forecast distribution. The ES has not been applied to electricity price forecasts so far. However, in the energy forecasting context is has been applied a few times, e.g. in [16] for wind power forecasting. The ES is built following [17] and then applied to our time series. We are going to pay more attention to this score because it takes into account dependency structures. As we mention above our contribution is to include correlation structures in the models as well as in the evaluation.

The remaining of the paper is organized as follows, in section 2 we explain the data, in section 3 we introduce the models, in section 4 we explain the estimation methods and how we generate the forecasts, in section 5 we describe the evaluation criteria, in section 6 we discuss the results and in section 7 we summarize our results and outline the most important facts.

2 Data description

As mentioned in the introduction we focus on off-peak and peak price series of the EPEX market, these series are based on German-Austrian day-ahead (hourly) electricity price. The peak series is calculated as the mean of the day-ahead price from 9^{th} hour of the day to the 20^{th} hour (12 hours in total), and consequently the off-peak price is the mean from 1^{st} to 8^{th} and from 21^{st} to 24^{th} hours of the day. The data starts in the 1^{st} of January 2014 and ends in the 31^{st} of December 2017. In order to calculate the higher moments and the dependencies, we follow the next notation,

$$\mathrm{m}_{i,j} = \mathbb{E}\left[\left(\frac{Y_{d,1} - \mu_{Y_{d,1}}}{\sigma_{Y_{d,1}}}\right)^i \left(\frac{Y_{d,2} - \mu_{Y_{d,2}}}{\sigma_{Y_{d,2}}}\right)^j\right],$$

where $Y_{d,1}$ and $Y_{d,2}$ refers to off-peak and peak time series with their means $\mu_{Y_{d,1}}$ and $\mu_{Y_{d,2}}$ and standard deviations $\sigma_{Y_{d,1}}$ and $\sigma_{Y_{d,2}}$. Below the descriptive statistics of both time series are shown,

	mean	sd	median	min	max	cor	skew	coskew	kurtosis
off-peak	28.30	8.74	29.36	-56.38	73.66	0.80	-1.61	-0.59	13.12
peak	35.48	13.79	35.09	-45.27	130.18	0.80	0.48	0.01	9.37

Where the number of days is 1461, sd is standard deviation. For the higher moments and dependencies cor makes reference to correlation, which is an estimator of $\mathrm{m}_{1,1}$ (same for both time series), skew to skewness with our notation estimated value of $\mathrm{m}_{3,0}$ and $\mathrm{m}_{0,3}$ for off-peak and peak, respectively. Similarly, coskew makes reference to the coskewness which are estimations of $\mathrm{m}_{2,1}$ and $\mathrm{m}_{1,2}$ and kurtosis refers to the estimated value of $\mathrm{m}_{4,0}$ and $\mathrm{m}_{0,4}$.

Table 1. Descriptive statistics of off-peak and peak prices

As expected, Table 1 shows that the mean and the standard deviation are higher in the peak time series. As the volatility is higher the range of peak is higher than the range of off-peak time series. The correlation shows a quite high positive linear relation between both time series. The skewness shows that the off-peak series is clearly asymmetric and that the peak series is slightly asymmetric. The coskewness coefficients show how are the variance of one time series and the mean of the other time series related, as it is observed in the Table 1 the relation between the off-peak central variance and the peak central mean is stronger compared to the relation the other way round, in the case of the $\mathrm{m}_{2,1} = -0,59$ means that the higher the value of the peak series the lower the variance of the off-peak one. Regarding the kurtosis in both cases the tails are heavier than the ones in the standard normal distribution which has kurstosis 3. After seen these results we can conclude that none of the time series follows a normal distribution.

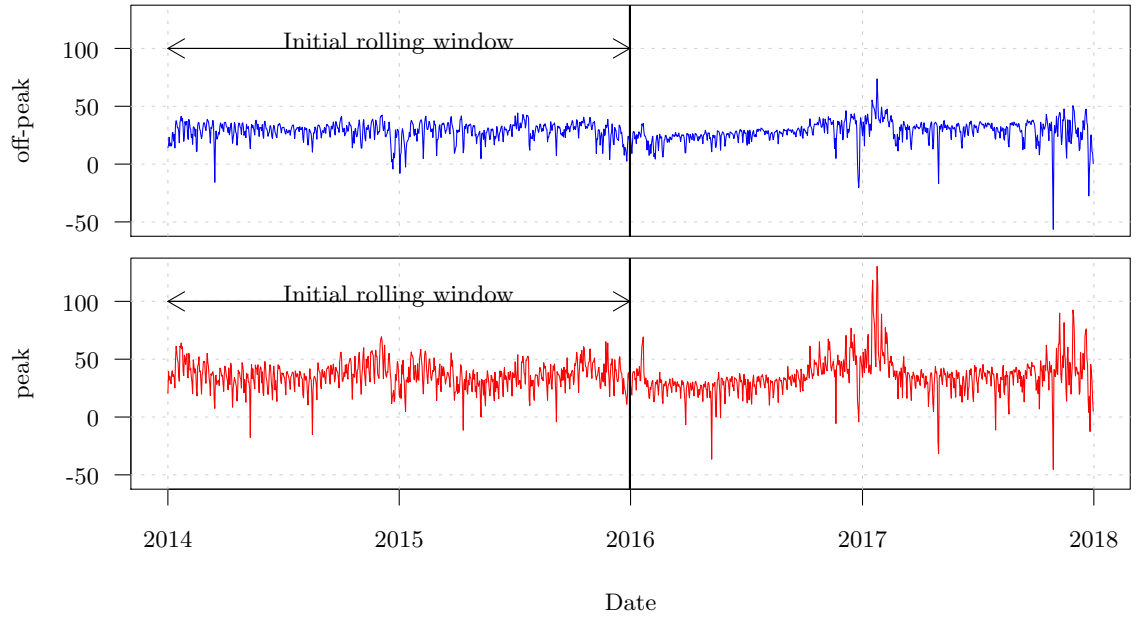
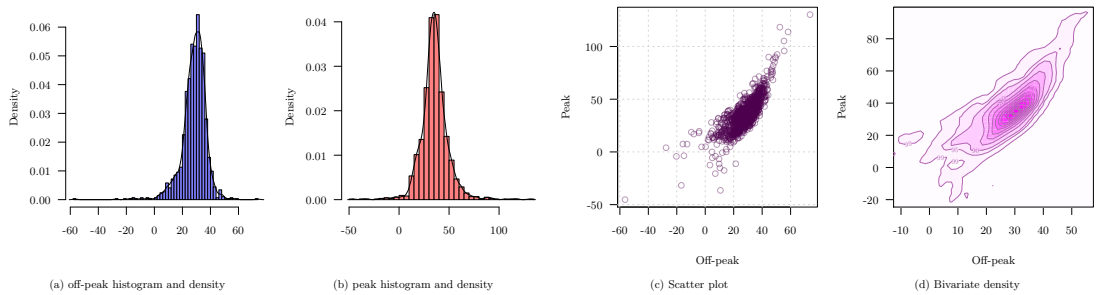


Fig. 1. off-peak and peak time series

In Figure 1 the off-peak and peak time series are depicted. The first two years are used only for estimation purpose and the last two years are first predicted and then used as observations of the following rolling windows, how the rolling windows are developed is explained in section 4. The Figure 1 is divided in two to underline this aspect. It is observed in the Figure 1 that at the beginning of the year 2017 the volatility was higher and so it was at the end of the year. As it can be observed in Table 1 and in the Figure 1 the volatility is higher, so is the mean comparing to those of the off-peak. However, generally the trend of the graphs is quite similar as shown by the correlation coefficient. In both cases there is evidence of volatility clustering and spikes.



(a) off-peak histogram and density (b) peak histogram and density (c) Scatter plot (d) Bivariate density

Fig. 2. Histograms and densities of off-peak and peak time series

The histograms and density functions of Figure 2 show the distribution of the two time series. As observed in the Table 1 both series have heavy tails and in the case of the off-peak the asymmetry is more pronounced. In the two cases there is evidence of spikes which are rare events where the price is extremely low or high. Regarding the scatter plot, strong correlation between both is checked which motivates us to include correlation structures in our models. Beside the Scatter plot it is possible to observe the bivariate density, which shows how is distributed the Scatter plot. In Figure 2d, the darker colors show higher quantiles of the distribution. The bivariate distribution confirms the intuition of the scatter plot where darker areas are those where more points are.

3 Models

The models we analyse in this paper are two step models. In the first step, we estimate the mean equation and in the second step we study the residuals from the previous step.

For simplification we define $\boldsymbol{Y}_d = (Y_{d,1}, Y_{d,2})'$ as the bivariate vector of the off-peak and peak prices, so that the index 1 corresponds to the off-peak price and 2 to the peak price.

3.1 ARX type models

In this subsection we introduce the conditional mean model we assume, the mean equation has been based on the mean models that [5] and [6] propose. In order to calculate the mean equation we assume autoregressive structure with exogenous variables (ARX) model for the peak and off-peak series. The ARX model is been proved to perform really good forecasting electricity prices in [5] and [6], we consider the both time series' mean model as,

$$\begin{aligned} Y_{d,i} =& \sum_{k=1}^{8} (\beta_{i,k,1} Y_{d-k,1} + \beta_{i,k,2} Y_{d-k,2}) \\ &+ \sum_{k=1}^{7} \left[(\beta_{i,k,3} + \beta_{i,k,4} Y_{d-1,1} + \beta_{i,k,5} Y_{d-1,2}) \, \mathrm{DoW}_d^k \right] + \epsilon_{d,i} \end{aligned} \tag{1}$$

where $i = 1, 2$ and DoW_d^j as day of the week dummy of day j at day d such that e.g. DoW_d^1 is equal to 1 if d falls on a Monday, $\mathrm{DoW}_d^2 = 2$ if d is on Tuesday etc. The residuals are $\epsilon_{d,1}$ and $\epsilon_{d,2}$, and by construction the mean of both error terms must be 0. The model has in total $p = 2 \times 8 + 3 \times 7 = 37$ parameters with corresponding parameter vector $\boldsymbol{\beta}$. Obviously model (1) is a linear model that can be written as

$$Y_{d,i} = \boldsymbol{X}_{d,i}' \boldsymbol{\beta}_i + \epsilon_{d,i} \tag{2}$$

where $\boldsymbol{X}_{d,i}$ and $\boldsymbol{\beta}_i$ are p-dimensional.

The error terms are considered to be distributed as,

$$\boldsymbol{\epsilon}_d \sim \mathcal{N}_2(\boldsymbol{0}, \, \boldsymbol{\Sigma}) \tag{3}$$

where $\boldsymbol{\epsilon}_d = (\epsilon_{d,1}, \epsilon_{d,2})'$, $\boldsymbol{0} = (0, 0)'$ and $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{\epsilon}_d$.

Model (1) covers already the major characteristics of electricity prices, esp. mean reverting properties, seasonal structure and cross-period effects. Just the volatility and price spikes are not captures by the assumed structure. Hence, for all the remaining models we consider the same mean equation, but modify the error model (3) to capture the missing effects.

3.2 ARX type models with independent jumps in the residuals

In this subsection we explain the model ARX-IJ, we consider mean MRJD in each of the residuals independently. This the standard OU process applied in electricity price forecasting several times as in [7] or widely discussed in [1]. Jump diffusion models are suitable to capture price spikes as we observe them in the tails of Figures 2a and 2b.

The model after Euler discretization is written as follows,

$$\begin{aligned} \boldsymbol{\epsilon}_d =& \boldsymbol{\epsilon}_{d,cont} + \boldsymbol{B}_d \boldsymbol{\epsilon}_{d,jump} \\ \boldsymbol{\epsilon}_{d,cont} \sim& \mathcal{N}_2(-\boldsymbol{\Lambda}\boldsymbol{\mu}, \, \boldsymbol{\Sigma}) \\ \boldsymbol{\epsilon}_{d,jump} \sim& \mathcal{N}_2(\boldsymbol{\mu}, \, \boldsymbol{\Gamma}) \end{aligned}$$

$$\text{where } \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$$\boldsymbol{B}_d = \begin{pmatrix} b_{d,1} & 0 \\ 0 & b_{d,2} \end{pmatrix} \text{ with } b_{d,i} \sim \mathrm{Ber}(\lambda_i) \text{ for } i = 1, 2, \text{and } \boldsymbol{\Gamma} = \begin{pmatrix} \gamma_1^2 & 0 \\ 0 & \gamma_2^2 \end{pmatrix},$$

where $\boldsymbol{\epsilon}_{d,cont}$ is continuous part of the error term and $\boldsymbol{\epsilon}_{d,jump}$ is the jump component. λ_i is the probability of jumps, Ber is the Bernoulli distribution, μ_i is the mean size of the jump, γ_i is the standard deviation of the jump and σ_i is the standard deviation of the continuous part all of them defined for $i = 1, 2$. We need the terms $-\lambda_1\mu_1$ and $-\lambda_2\mu_2$ in the continuous term in order to ensure that the mean of $\epsilon_{d,1}$ and $\epsilon_{d,2}$ and is equal to 0, as it must be by construction. In this model we assume that the Bernoulli random variables $b_{d,1}$ and $b_{d,2}$ are independent. The conditional error term $\boldsymbol{\epsilon}_d | \boldsymbol{B}_d = \boldsymbol{\epsilon}_{d,cont} + \boldsymbol{B}_d \boldsymbol{\epsilon}_{d,jump} | \boldsymbol{B}_d$ is distributed as follows,

$$\boldsymbol{\epsilon}_d | \boldsymbol{B}_d \sim \mathcal{N}_2(\boldsymbol{0}, \, \boldsymbol{\Sigma} + \boldsymbol{B}_d \boldsymbol{\Gamma} \boldsymbol{B}_d'). \tag{4}$$

For the unconditional distribution of $\boldsymbol{\epsilon}_d$ we first notice that with $\text{Var}[XY] = \mathbb{E}[X]^2\text{Var}[Y] + \text{Var}[X]\mathbb{E}[Y]^2 + \text{Var}[X]\text{Var}[Y]$ it holds

$$\begin{aligned}\text{Var}[b_i\epsilon_{d,jump,i}] &= \mathbb{E}[b_i]^2\text{Var}[\epsilon_{d,jump,i}] + \text{Var}[b_i]\mathbb{E}[\epsilon_{d,jump,i}]^2 + \text{Var}[b_i]\text{Var}[\epsilon_{d,jump,i}] \\ &= \lambda_i^2\gamma_i^2 + \lambda_i(1-\lambda_i)\mu_i^2 + \lambda_i(1-\lambda_i)\gamma_i^2 = \lambda_i((1-\lambda_i)\mu_i^2 + \gamma_i^2). \end{aligned} \tag{5}$$

Thus by the independence of all occurring random variables it holds,

$$\text{Var}[\boldsymbol{\epsilon}_d] = \boldsymbol{\Sigma} + \text{Var}[\boldsymbol{B}_d\boldsymbol{\epsilon}_{d,jump}] = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}((\boldsymbol{I}-\boldsymbol{\Lambda})\text{Diag}(\boldsymbol{\mu})^2 + \text{Diag}(\boldsymbol{\Gamma})) \tag{6}$$

as $\text{Cov}[b_{d,1}\epsilon_{d,jump,1}, b_{d,2}\epsilon_{d,jump,2}] = 0$. However, it is clear that $\boldsymbol{\epsilon}_d$ does not follow a bivariate normal distribution.

3.3 ARX type models with bivariate jumps in the residuals

The next model we introduce, the ARX-BiJ model, is somehow related to the previous one as it is based on MRJD structure but in this case the jump component is assumed to be bivariate, bivariate Bernoulli to be more precise. This dependency structure in the jump occurrence is one of the contributions to the existing literature. Moreover, we assume additionally that the jump sizes can be correlated as well. We write the model as,

$$\begin{aligned}\boldsymbol{\epsilon}_d =& \boldsymbol{\epsilon}_{d,cont} + \boldsymbol{B}_d\boldsymbol{\epsilon}_{d,jump} \quad \text{with} \quad \boldsymbol{\epsilon}_{d,cont} \sim \mathcal{N}_2(-\boldsymbol{\Lambda}\boldsymbol{\mu},\, \boldsymbol{\Sigma}) \quad \text{and} \quad \boldsymbol{\epsilon}_{d,jump} \sim \mathcal{N}_2(\boldsymbol{\mu},\, \boldsymbol{\Gamma}), \\ &\text{where } \boldsymbol{\mu} = \begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}, \boldsymbol{\Gamma} = \begin{pmatrix}\gamma_1^2 & \varrho\gamma_1\gamma_2 \\ \varrho\gamma_1\gamma_2 & \gamma_2^2\end{pmatrix} \\ &\boldsymbol{B}_d = \begin{pmatrix}b_{d,1} & 0 \\ 0 & b_{d,2}\end{pmatrix} \quad \text{with } \text{diag}(\boldsymbol{B}_d) \sim \mathbf{Ber}_2(\boldsymbol{P}) \\ &\text{with probabilities } \boldsymbol{P} = \begin{pmatrix}p_{0,0} & p_{1,0} \\ p_{0,1} & p_{1,1}\end{pmatrix}, \boldsymbol{\Lambda} = \begin{pmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{pmatrix} = \begin{pmatrix}p_{1,0}+p_{1,1} \\ p_{0,1}+p_{1,1}\end{pmatrix}\end{aligned}$$

As mentioned, now we assume that the jumps arrivals are bivariate $\mathbf{Ber}_2(\boldsymbol{P})$ distributed with probabilities $\boldsymbol{P}$. In this case, $p_{0,0}$ is probability of no jump, $p_{1,0}$ is the probability of jump occurrence only in the off-peak component, $p_{1,1}$ is the probability of having a jump in both components at the same time and $p_{0,1}$ is the probability of jump only in the peak component. Therefore, $\lambda_2 = p_{0,1} + p_{1,1}$ is the total probability of jump in the peak component, similarly $\lambda_1 = p_{1,0} + p_{1,1}$ is the probability of jump in off-peak component. The condition $p_{1,0} + p_{1,1} + p_{0,1} + p_{0,0} = 1$ must be hold. Unlike the previous model in this case $b_{d,1}$ and $b_{d,2}$ are not independent, they have to coincide in no jump with probability $p_{0,0}$ and in jump with probability $p_{1,1}$. For the bivariate Bernoulli setting we follow [11]. Assuming bivariate jump process we capture dependency structure in the continuous component as well as in the jump component.

3.4 ARX type models with bivariate jumps in the residuals with no constant mean.

In this subsection the ARX-BiJ-μ_d model is presented, this model is very similar to the previous one but in this case we assume that the mean of the jump depends on the previous observed price. In order not to make to tedious to the reader we only note the things that change from the previous model, that is,

$$\begin{aligned}&\boldsymbol{\mu}_d = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1\boldsymbol{Y}_{d-1} \text{ and } \boldsymbol{\epsilon}_d = \boldsymbol{\epsilon}_{d,cont} + \boldsymbol{B}_d\boldsymbol{\epsilon}_{d,jump} \\ &\text{with } \boldsymbol{\epsilon}_{d,cont} \sim \mathcal{N}_2(-\boldsymbol{\Lambda}\boldsymbol{\mu}_d,\, \boldsymbol{\Sigma}), \text{diag}(\boldsymbol{B}_d) \sim \mathbf{Ber}_2(\boldsymbol{P}) \text{ and } \boldsymbol{\epsilon}_{d,jump} \sim \mathcal{N}_2(\boldsymbol{\mu}_d,\, \boldsymbol{\Gamma}),\end{aligned} \tag{7}$$

In this model we try to capture the effect of the previous price in the mean of the jump component.

3.5 ARX type models with CCC-GARCH

In the next model (ARX-GARCH) we consider bivariate constant conditional correlation GARCH (CCC-GARCH) structures, first introduced by [18]. We follow [19] in the implementation,

$$\begin{aligned}\epsilon_{d,i} =& \sigma_{d,i}z_{d,i} \\ \sigma_{d,i}^2 =& \alpha_{0,i} + \alpha_{1,i}\epsilon_{d-1,i}^2 + \alpha_{2,i}\sigma_{d-1,i}^2 \text{ for } i = 1, 2, \\ \boldsymbol{\epsilon}_d \sim& \mathcal{N}_2(\mathbf{0},\, \boldsymbol{\Sigma}_d) \quad \text{where} \quad \boldsymbol{\Sigma}_d = \begin{pmatrix}\sigma_{d,1}^2 & \rho\sigma_{d,1}\sigma_{d,2} \\ \rho\sigma_{d,1}\sigma_{d,2} & \sigma_{d,2}^2\end{pmatrix}\end{aligned}$$

where $z_{d,1}$ and $z_{d,2}$ are independent white noises with standard deviation of 1. The parameters of the GARCH structure must fulfill $\alpha_{0,i}, \alpha_{1,i}, \alpha_{2,i} > 0$ and $\alpha_{1,i} + \alpha_{2,i} < 1$ conditions in order the time series to be stationary. In this model we assume that the correlation is constant between the two time series and no cross-dependencies between the volatility series. This type of structures are often used in the literature to forecast multivariate time series, for instance in [20] and [21] apply CCC-GARCH models in electricity markets.

3.6 ARX type models with bivariate jumps in the residuals with no constant mean and CCC-GARCH

In the last model , the ARX-BiJ-μ_d-GARCH model, we include CCC-GARCH structures to the continuous component of the model described in Equation (7). This is the most complex model and it is the combination of the models ARX-BiJ-μ_d and ARX-GARCH,

$$\begin{aligned}
&\boldsymbol{\mu}_d = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 \boldsymbol{Y}_{d-1} \text{ and } \boldsymbol{\epsilon}_d = \boldsymbol{\epsilon}_{d,cont} + \boldsymbol{B}_d \boldsymbol{\epsilon}_{d,jump} \\
&\text{with } \boldsymbol{\epsilon}_{d,cont} \sim \mathcal{N}_2(-\boldsymbol{\Lambda}\boldsymbol{\mu}_d,\, \boldsymbol{\Sigma}_d), \operatorname{diag}(\boldsymbol{B}_d) \sim \mathbf{Ber}_2(\boldsymbol{P}) \text{ and } \boldsymbol{\epsilon}_{d,jump} \sim \mathcal{N}_2(\boldsymbol{\mu}_d,\, \boldsymbol{\Gamma}), \\
&\sigma^2_{d,i} = \alpha_{0,i} + \alpha_{1,i}\epsilon^2_{d-1,i} + \alpha_{2,i}\sigma^2_{d-1,i} \text{ for } i = 1, 2, \\
&\text{where } \ \boldsymbol{\Sigma}_d = \begin{pmatrix} \sigma^2_{d,1} & \rho\sigma_{d,1}\sigma_{d,2} \\ \rho\sigma_{d,1}\sigma_{d,2} & \sigma^2_{d,2} \end{pmatrix}
\end{aligned}$$

where all the components are assumed to be distributed as in the previous subsections. This model as it is the most complex it has the highest amount of parameters to estimate. The model is able to capture all the aspects mentioned above.

4 Estimation and Forecasting

For the estimation we assume that we have D observations available. We denote the resulting price vectors and regression matrix are $\mathbb{Y}_i = (Y_{1,i}, , \ldots, Y_{D,i})'$ and $\mathbb{X}_i = (\boldsymbol{X}'_{1,i}, , \ldots, \boldsymbol{X}'_{D,i})'$ that correspond to the regression equation (2).

For the estimation of the ARX model (Equation (2)) we apply two different estimation methods OLS[3] and elastic net. As we get two different estimations and therefore two different forecasts we note them as ARX-OLS and as ARX-enet, respectively.

Using the OLS estimator the estiamted values are,

$$\widehat{\boldsymbol{\beta}}_i^{\text{OLS}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \left[\|\mathbb{Y}_i - \mathbb{X}_i'\boldsymbol{\beta}\|_2^2 \right],$$

The second estimation method applied to estimate Equation (1) is the elastic net, introduced by [22], which is very similar to OLS but it has quadratic and linear penalties. However, for defining the elastic net estimator it is crucial to consider the corresponding scaled OLS problem. Hence, we introduce $\widetilde{\mathbb{Y}}_i$ and $\widetilde{\mathbb{X}}_i$ as scaled response vector and scaled regression matrix. We require them scaled in such a way that any column has a zero mean and standard deviation of 1.

Given the scaled OLS problem the scaled elastic net estimator is given by the optimization problem,

$$\widetilde{\widehat{\boldsymbol{\beta}}}_i^{\text{enet}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \left[\|\widetilde{\mathbb{Y}}_i - \widetilde{\mathbb{X}}_i'\boldsymbol{\beta}\|_2^2 + \lambda \left(\frac{1-\alpha}{2} ||\boldsymbol{\beta}||_2^2 + \alpha ||\boldsymbol{\beta}||_1 \right) \right],$$

λ and α are tuning parameters that characterize the penalty term $\lambda\left(\frac{1-\alpha}{2}||\boldsymbol{\beta}||_2^2 + \alpha||\boldsymbol{\beta}||_1\right)$. We receive the (unscaled) elastic net estimator $\widehat{\boldsymbol{\beta}}_i^{\text{enet}}$ simply by rescaling of $\widetilde{\widehat{\boldsymbol{\beta}}}_i^{\text{enet}}$. If $\alpha = 1$ the estimation method is equal to the lasso penalty, developed by [23] and when $\alpha = 0$ is equal to the ridge penalty, first introduced by [24]. The lasso estimator exhibits the sparsity property which means that for certain values of λ the resulting solution sets irrelevant parameters to zero while keeping relevant parameters non-zero. The lasso estimation enjoys some popularity in electricity price forecasting, see [4], [2], [25] and [26], among others.

The elastic net can be seen as an augmented data lasso shrinkage with some ridge elements. Similarly to the lasso, the elastic net has automatic sparsity property for $\alpha \in (0, 1)$. Lasso method does not perform good with highly correlated variables that is why we include the ridge penalty. The elastic net is applied by [5] in the electricity price forecasting context. Their findings suggest

[3] In order to avoid perfect collinearity in the OLS estimation we drop the interaction between Wednesday and the previous observations for both time series.

that $\alpha = 0.5$ is a good choice for applications, that we will apply here as well. We choose λ by 10-fold cross validation.

We compare the results for both estimation techniques and as the elastic net provides better forecasting results, for the estimation of the second step the residuals are calculated by elastic net. In the second step residuals are estimated by maximum likelihood[4] using Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. This algorithm is a quasi-Newton method for non-linear optimization. In order to apply maximum likelihood estimation we need a log-likelihood function which changes depending on the structure of the residuals we assume. The log-likelihood function is based on the assumed distribution for each one of the different models explained in section 3.

Once estimated all the parameters, off-peak and peak time series are simulated. In our case we simulate $H = 7$ horizons and for each of the horizons $M = 16000$ paths are generated. All the simulations (sometimes called ensemble) can be seen as multivariate probabilistic forecasts as they approximate well the underlying distribution of the forecasts. All relevant properties can be derived from these paths, it is possible to analyse the only marginal properties of each one of the predicted horizons. The estimation and simulation process is repeated $N = 731$ times, as mentioned in section 2 by a rolling window. The first estimation is made using the first two years of the data, then the next $H = 7$ days are predicted with M paths in each one of the horizons. Afterwards, the estimation sample is shifted one day forward and the process is repeated.

5 Evaluation criteria

In this section we introduce the evaluation criteria we use to evaluate probabilistic forecasting. We use in total four different criteria: MAE, MSE, PB and ES. After explaining the scores we use, we briefly introduce the Diebold-Mariano (DM) test which we use to test whether the difference, taking the models in pairs, in the forecasting performance are significant or not. In order to compute the DM test it is necessary to define a loss function this is the reason we introduce each criteria with the corresponding loss function.

For the evaluation of the point forecasts we consider the popular measures MAE and MSE. The MAE is a strictly proper forecasting criterion for the median and the MSE a strictly proper evaluation criterion for the mean. Here, strictly proper means that only the perfect model minimizes the corresponding criterion. Therefore we define $\widehat{Y}^{\text{med}}_{d,i}$ as a median forecasts and $\widehat{Y}_{d,i}$ mean for day d and volatility series i derived from the sample counterparts of the M simulated paths. The MAE and MSE are defined using the absolute error (AE) and the squared error (SE). Thus, with

$$\text{AE}_{d,i} = |Y_{d,i} - \widehat{Y}^{\text{med}}_{d,i}| \tag{8}$$

$$\text{SE}_{d,i} = \left(Y_{d,i} - \widehat{Y}_{d,i}\right)^2 \tag{9}$$

we define

$$\text{MAE}_i = \frac{1}{N}\sum_{d=1}^{N} \text{AE}_{d,i} \tag{10}$$

$$\text{MSE}_i = \frac{1}{N}\sum_{d=1}^{N} \text{SE}_{d,i} \tag{11}$$

for $i = 1, 2$. Hence, we can evaluate the point forecasts for the off-peak and peak price separately.

The two criteria introduced above are the most used ones in the literature but we are more interested in the marginal properties of the models and to that purpose we use the PB (PB).

The PB measures the distance on each one of the quantiles. Therefore, similarly as for the median and mean forecasts, we define $\widehat{Y}_{d,q,i}$ as a forecasts for the q-quantile day d and time series i. We get these quantile forecasts by taking the sample quantile of our M simulated paths. The pinball loss function is computed as follows,

$$PB_{d,q,i} = \begin{cases} (1-q)(\widehat{Y}_{d,q,i} - Y_{d,i}), & \text{if } \widehat{Y}_{d,q,i} \geq Y_{d,i} \\ q(Y_{d,i} - \widehat{Y}_{d,q,i}), & \text{if } \widehat{Y}_{d,q,i} < Y_{d,i} \end{cases} \quad \text{for} \quad i = 1, 2, \tag{12}$$

where $q \in \boldsymbol{Q}$ is a quantile, in our case $\boldsymbol{Q} = \{Q_q\}_{q\in\{1,\ldots,K\}} = \{0.01, 0.02, \ldots, 0.99\}$ with $K = 99$. $\widehat{Y}_{d,q,i}$ stands for the estimated price of quantile q in day d and time series i and $Y_{d,i}$ is the

[4] In the ARX-IJ model ρ is estimated in a third stage as $\rho = \text{Cor}(\epsilon_{d,1}, \epsilon_{d,2})$

observed value at day d and time series i. If the objective is to calculate the PB of quantile q, series i and N days it is computed as follows,

$$PB_{q,i} = \frac{1}{N}\sum_{d=1}^{N} PB_{d,q,i} \quad \text{for} \quad i = 1,2 \quad \text{and} \quad q \in Q,$$

Thus the PB for N days it is computed by averaging across the $\boldsymbol{Q}$ quantiles,

$$PB_{i} = \frac{1}{K}\sum_{q=1}^{K} PB_{Q_q,i} \quad \text{for} \quad i = 1,2 \quad \text{and} \quad q \in Q,$$

where K is the number of quantiles. Note that if the distance in the quantile grid $\boldsymbol{Q}$ converges to 0 then the PB converges to the probabilistic forecasting evaluation measure CRPS (continuous ranked probability score) which is strictly proper with respect to the (marginal) distribution of Y_i. For further information on the PB score see [27].

When applying PB we can observe the whole marginal properties of the forecasts, as we can observe the forecasting performance of the different models in each one of the quantiles. Therefore, the first criterion introduced in this section is nothing but an special case of PB when $q = 0.5$. With this score we are able to compare how the different models capture spikes, as we can analyse the behaviour in the tails.

The last criterion we use is the ES which is a generalization of the CRPS. The ES is the only score that takes into account the dependency structures, this score is applied to all the variables at the same time in order to take into the correlation. In both cases for the modelling and the evaluation we pay a lot attention to the dependency structure as this is main contribution of this paper. The loss function of the is computed as follows,

$$\begin{aligned} \mathrm{ES}_d =& \mathrm{ED}_d + \frac{1}{2}\mathrm{EI}_d \\ \mathrm{ED}_d =& \frac{1}{M}\sum_{m=1}^{M} \left\| \boldsymbol{Y}_d^{[m]} - \boldsymbol{Y}_d \right\|_2 \\ \mathrm{EI}_d =& \frac{1}{M}\sum_{m=1}^{M} \left\| \boldsymbol{Y}_d^{[m]} - \boldsymbol{Y}_d^{[m+1]} \right\|_2 \quad \text{where} \quad \boldsymbol{Y}_d^{[M+1]} = \boldsymbol{Y}_d^{[1]} \end{aligned} \tag{13}$$

$\boldsymbol{Y}_d^{[m]}$ for $m = 1, \ldots, M$ is the predicted m^{th} path of the multivariate data for day d. Our estimator for the ES is based on [17]. Equation (13) is the ES for day d, thus to compute the ES for N days it is calculated as,

$$\mathrm{ES} = \frac{1}{N}\sum_{d=1}^{N} \mathrm{ES}_d.$$

The next step is to check whether the difference among the forecasting performances with each criteria are significantly different from zero or not. For the significance test we use the DM test which compares models in pairs. As mentioned above we need a loss function to apply the DM test, loss functions as the ones written above. Let L_d denote the loss function of a certain model; the loss differential between models $\mathbb{A}$ and $\mathbb{B}$ is defined as $\delta_{d,\mathbb{A},\mathbb{B}} = L_{d,\mathbb{A}} - L_{d,\mathbb{B}}$. The only assumption we need is that the loss differential has to be covariance stationary. In order to apply the DM test we compute,

$$\frac{\widetilde{\delta}_{\mathbb{A},\mathbb{B}}}{\sigma_{\widetilde{\delta}_{\mathbb{A},\mathbb{B}}}} \sim \mathcal{N}_1(0,1)$$

where $\widetilde{\delta}_{\mathbb{A},\mathbb{B}} = \frac{1}{N}\sum_{d=1}^{N} \delta_{d,\mathbb{A},\mathbb{B}}$ and $\sigma_{\widetilde{\delta}_{\mathbb{A},\mathbb{B}}}$ is the standard error which we estimate by the corresponding sample counterpart. Further information on the DM test in [28] and [29].

6 Results

In this section we evaluate the forecasting performance of each one of the models. For that purpose the criteria introduced in section 5 are used. Afterwards, the DM test is applied in order to check whether the differences among the models in pairs are significant or not[5] for each one

[5] DM test results for all criteria and all horizons are available upon request.

of the criteria. The forecasting horizon (H) is 7 which means that for each one of the rolling windows we get the predictions for the following 7 days. We denominate those horizon as H1, ..., H7 and the models are the ones explained in section 3. In order to make the tables more visual we use different colors for the forecasting performance. The greener is the color the smaller is the error in each one of the horizons and the closer the color is from red the poorer is the performance. Models 1 to 7 are models ARX-OLS, ARX-enet, ARX-IJ, ARX-BiJ, ARX-BiJ-μ_d, ARX-GARCH and ARX-BiJ-μ_d-GARCH, respectively.

off-peak	H1	H2	H3	H4	H5	H6	H7
1	4.117 (0.192)	5.519 (0.257)	5.92 (0.262)	6.112 (0.271)	6.215 (0.274)	6.284 (0.273)	6.321 (0.271)
2	4.018 (0.193)	5.416 (0.253)	5.83 (0.26)	6.037 (0.267)	6.118 (0.27)	6.193 (0.27)	6.252 (0.269)
3	4.011 (0.193)	5.423 (0.254)	5.825 (0.262)	6.034 (0.268)	6.111 (0.271)	6.198 (0.271)	6.253 (0.269)
4	4.03 (0.196)	5.419 (0.256)	5.795 (0.264)	6.017 (0.271)	6.118 (0.276)	6.222 (0.275)	6.271 (0.274)
5	4.033 (0.196)	5.429 (0.255)	5.806 (0.263)	6.041 (0.271)	6.113 (0.275)	6.224 (0.274)	6.27 (0.273)
6	4.021 (0.195)	5.418 (0.253)	5.823 (0.26)	6.029 (0.267)	6.121 (0.271)	6.196 (0.27)	6.251 (0.269)
7	4.049 (0.195)	5.46 (0.256)	5.876 (0.264)	6.119 (0.271)	6.199 (0.275)	6.319 (0.273)	6.344 (0.274)

peak	H1	H2	H3	H4	H5	H6	H7
1	6.713 (0.3)	7.966 (0.344)	8.259 (0.351)	8.395 (0.364)	8.447 (0.37)	8.494 (0.372)	8.543 (0.374)
2	6.542 (0.292)	7.825 (0.34)	8.136 (0.354)	8.272 (0.365)	8.344 (0.371)	8.398 (0.374)	8.49 (0.377)
3	6.522 (0.292)	7.843 (0.34)	8.128 (0.353)	8.268 (0.363)	8.34 (0.37)	8.408 (0.372)	8.501 (0.375)
4	6.537 (0.292)	7.842 (0.338)	8.157 (0.351)	8.296 (0.361)	8.405 (0.368)	8.478 (0.37)	8.599 (0.372)
5	6.531 (0.292)	7.848 (0.34)	8.123 (0.353)	8.258 (0.363)	8.325 (0.371)	8.395 (0.372)	8.486 (0.374)
6	6.515 (0.292)	7.817 (0.34)	8.12 (0.354)	8.272 (0.364)	8.344 (0.37)	8.416 (0.373)	8.495 (0.376)
7	6.581 (0.294)	7.873 (0.34)	8.231 (0.352)	8.384 (0.359)	8.506 (0.366)	8.595 (0.367)	8.694 (0.371)

The main number is the mean of the MAE over the 731 rolling windows for both time series. The number in brackets shows the corresponding standard deviation.

Table 2. MAE

Table 2 shows the results for the MAE criterion for the two time series. As explained above this is a special case of the PB and it compares the performance on the median. As it is observed in the Table 2 the differences among the models are not too big, the only clear result is that the performance of the ARX-OLS model is significantly poorer. Overall, we can say that results for the ARX-GARCH and ARX-IJ are slightly better but as mentioned the forecasts are not significantly better, the DM test confirms this fact. It is important to underline that when comparing the OLS and elastic net estimation methods the forecasting results are significantly better using the later one.

off-peak	H1	H2	H3	H4	H5	H6	H7
1	65.32 (7.26)	115.07 (11.88)	128.61 (13.1)	137.97 (13.33)	142.41 (13.55)	144.02 (13.42)	144.28 (13.24)
2	65.66 (7.41)	111.75 (12.05)	124.43 (12.78)	132.63 (13.23)	135.88 (13.33)	137.52 (13.23)	138.4 (13.09)
3	65.55 (7.38)	110.53 (12.04)	123.18 (12.86)	130.73 (13.18)	133.92 (13.3)	135.59 (13.2)	136.44 (13.06)
4	69.55 (7.71)	115.73 (12.11)	130.15 (12.98)	139.57 (13.31)	144.55 (13.47)	147.27 (13.31)	149.04 (13.21)
5	69.42 (7.75)	115.58 (12.16)	129.69 (12.92)	139.07 (13.32)	143.7 (13.47)	146.78 (13.32)	149.27 (13.18)
6	66.34 (7.68)	111.84 (12.02)	125.11 (12.89)	132.88 (13.2)	136.93 (13.37)	138.82 (13.23)	139.74 (13.11)
7	68.37 (7.64)	115.84 (12.2)	131.71 (12.92)	142.74 (13.36)	149.05 (13.42)	154.46 (13.3)	159.47 (13.22)

peak	H1	H2	H3	H4	H5	H6	H7
1	162.56 (14.32)	221.05 (17.02)	239.02 (17.72)	253.97 (18.75)	261.68 (19.18)	264.97 (19.15)	268.75 (19.26)
2	158.29 (13.23)	216.83 (16.66)	236.83 (17.82)	249.23 (18.72)	256.19 (18.9)	260.04 (19.24)	264.58 (19.33)
3	157.89 (13.18)	214.31 (16.66)	232.11 (17.8)	242.94 (18.52)	249.74 (18.87)	253.17 (19.04)	258.14 (19.09)
4	176.66 (13.39)	236.05 (16.77)	256.09 (17.91)	268.99 (18.69)	278.19 (19.2)	283.17 (19.23)	290 (19.28)
5	174.97 (13.45)	233.05 (16.86)	250.93 (18.01)	262.24 (18.79)	269.6 (19.37)	273.2 (19.28)	278.22 (19.23)
6	158.23 (13.31)	218.01 (16.7)	239.71 (17.9)	253.37 (18.68)	261.71 (19)	266.64 (19.09)	272.81 (19.27)
7	179.11 (13.81)	244.84 (17.2)	272.05 (18.44)	292.09 (19.25)	308.64 (19.9)	320.4 (20.47)	337.77 (21.53)

The main number is the mean of the MSE over the 731 rolling windows for both time series. The number in brackets shows the corresponding standard deviation.

Table 3. MSE

In Table 3 we show the forecasting performance in the mean. After looking to the Table 3 and the DM test, it looks like bivariate jump structures are not effective to capture the mean behaviour as the forecasting performance is poor according to the MSE criterion. At the same time, this Table shows that the forecasts of the ARX-IJ are better than other models and the DM test confirms that the differences are significant, with the exception of the H1 where the difference among ARX-IJ and ARX-enet is not significant for both time series. This fact means that for the mean forecast it is important to introduce jump structures but the dependency structures are not that relevant. The superiority of the ARX-IJ forecasting performance is more pronounced when the horizons are increased. According to this criterion the elastic net forecasts are significantly better than the OLS forecasts when simple ARX model is simulated.

As showed in the previous section the PB takes into account the whole distribution of the forecast paths quantile by quantile. As showed in Table 4 in the case of the peak time series the ARX-BiJ-μ_d models forecasts outperforms the other models except from one case, as in H2

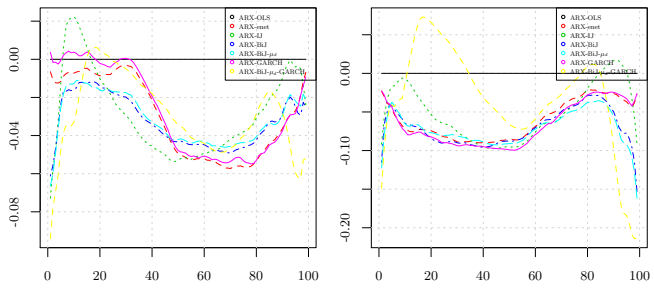
off-peak	H1	H2	H3	H4	H5	H6	H7
1	1.582 (0.01)	2.119 (0.013)	2.249 (0.013)	2.32 (0.013)	2.359 (0.014)	2.381 (0.013)	2.379 (0.013)
2	1.552 (0.01)	2.073 (0.013)	2.218 (0.013)	2.288 (0.013)	2.322 (0.013)	2.349 (0.013)	2.357 (0.013)
3	1.553 (0.01)	2.082 (0.013)	2.227 (0.013)	2.294 (0.014)	2.333 (0.014)	2.359 (0.014)	2.367 (0.014)
4	1.549 (0.01)	2.072 (0.013)	2.21 (0.013)	2.284 (0.014)	2.323 (0.014)	2.353 (0.014)	2.362 (0.014)
5	1.55 (0.01)	2.074 (0.013)	2.215 (0.013)	2.291 (0.014)	2.323 (0.014)	2.353 (0.014)	2.364 (0.014)
6	1.555 (0.01)	2.074 (0.013)	2.218 (0.013)	2.289 (0.013)	2.33 (0.014)	2.356 (0.014)	2.366 (0.014)
7	1.552 (0.01)	2.095 (0.013)	2.249 (0.013)	2.329 (0.014)	2.368 (0.014)	2.406 (0.014)	2.417 (0.014)

peak	H1	H2	H3	H4	H5	H6	H7
1	2.565 (0.015)	3.046 (0.017)	3.153 (0.017)	3.22 (0.018)	3.26 (0.018)	3.278 (0.018)	3.298 (0.018)
2	2.504 (0.014)	2.991 (0.017)	3.124 (0.017)	3.188 (0.018)	3.234 (0.018)	3.257 (0.018)	3.294 (0.018)
3	2.513 (0.015)	3.01 (0.017)	3.13 (0.018)	3.189 (0.018)	3.234 (0.019)	3.26 (0.019)	3.298 (0.019)
4	2.494 (0.015)	2.985 (0.017)	3.118 (0.018)	3.177 (0.018)	3.23 (0.018)	3.259 (0.018)	3.303 (0.019)
5	2.492 (0.015)	2.988 (0.017)	3.108 (0.018)	3.169 (0.018)	3.208 (0.018)	3.239 (0.019)	3.273 (0.019)
6	2.498 (0.014)	2.987 (0.017)	3.119 (0.017)	3.184 (0.018)	3.229 (0.018)	3.256 (0.018)	3.292 (0.018)
7	2.532 (0.015)	3.053 (0.017)	3.195 (0.018)	3.269 (0.019)	3.328 (0.019)	3.361 (0.019)	3.397 (0.019)

The main number is the mean of the Pinball score within the 99 quantiles for each one of the 731 rolling windows. The number in brackets shows the corresponding standard deviation.

Table 4. Pinball score

the PB score of the ARX-BiJ model is lower. On the other hand, in the case of the off-peak we see that the for the first horizons the ARX-BiJ forecasts are the best ones but after the 6^{th} horizon the ARX-enet makes the best forecasting performance. This fact is curious because the ARX-enet does not take into account heavy tails and it is expected that a models with jumps would capture the tail behaviour more efficiently. We must underline though that the differences among the models are not significant according to the DM test. The DM test only concludes that for both time series the forecasting performance is significantly poorer for the ARX-OLS and the ARX-BiJ-μ_d-GARCH models. As with the MAE and MSE criteria with the PB score the elastic net estimation method provides better forecasts than the OLS.



(a) off-peak quantile performance (b) peak quantile performance

Fig. 3. Performance on each quantile for first horizon

The Figure 3 shows $PB(model) - PB(OLS)$ for the first predicted horizon quantile by quantile, therefore all the models are compared to the ARX-OLS. If we focus on the model ARX-BiJ-μ_d-GARCH, which is the most complex one, we observe that in the first and the last quantiles it is the best model. However, in the middle quantiles it has the worst forecasting results, it is even more pronounced for the peak time series. The behaviour of the forecasts with models ARX-BiJ-μ_d and ARX-BiJ is similar in both cases. Regarding the ARX-enet in the off-peak time series it is observed that the in the middle quantiles the forecasting performance is the best together with the ARX-GARCH.

So far we have distinguished between the two time series because dependencies are not taken into account in the criteria mentioned above. It is observed for the three previous criteria that the off-peak time series has a lower error term which means that forecasts are more accurate according to the three criteria. The next score takes into account dependency structures which are the key feature of this paper.

Table 5 shows the ES for our seven models and our seven horizons. It is observed that the best forecasts are made by the ARX-BiJ-μ_d which is sensible taking into account in this forecasting model takes into account dependency structure. We would expect that the ARX-BiJ-μ_d-GARCH capturing all features of the time series better but it may the case that as there are to many parameters to estimate with our starting values we have hit a local maximum[6].

[6] Other starting values could be used but it would increase a lot the computational time and, as 731 estimations are made in total, increasing the number of starting values would have made estimation infeasible.

Values	H1	H2	H3	H4	H5	H6	H7
1	731.46 (24.12)	932.9 (29.28)	1000.6 (30.09)	1036.75 (31.26)	1056.79 (31.56)	1069.21 (31.46)	1078.45 (31.29)
2	731.44 (23.85)	922.89 (29.43)	988.47 (30.41)	1020.57 (31.61)	1038.11 (31.9)	1050.09 (31.93)	1062.48 (31.98)
3	638.33 (23.72)	806.99 (29.32)	864.44 (30.4)	891.4 (31.4)	907.72 (31.72)	919.05 (31.72)	928.85 (31.71)
4	588.11 (23.84)	747.49 (29.04)	807.22 (30)	839.12 (30.9)	860.91 (31.26)	875.71 (31.2)	887.44 (31.13)
5	583.19 (23.92)	735.43 (29.14)	789.13 (30.2)	816.7 (31.15)	833.62 (31.59)	845.63 (31.47)	855.38 (31.3)
6	729.66 (24.05)	917.93 (29.31)	983.39 (30.39)	1013.91 (31.28)	1033.36 (31.64)	1046.18 (31.61)	1058.63 (31.56)
7	614.33 (24.03)	786.33 (29.34)	855.12 (30.07)	895.34 (30.85)	924.59 (31.14)	948.82 (31.06)	968.25 (31.17)

DM-test	1	2	3	4	5	6	7
1		0 (0.502)	25.02 (>0.999)	32.58 (>0.999)	34.83 (>0.999)	0.47 (0.68)	24.39 (>0.999)
2	0 (0.498)		54.72 (>0.999)	61.92 (>0.999)	66.52 (>0.999)	1.71 (0.956)	36.46 (>0.999)
3	−25.02 (<0.001)	−54.72 (<0.001)		30.9 (>0.999)	34.53 (>0.999)	−50.4 (<0.001)	8.48 (>0.999)
4	−32.58 (<0.001)	−61.92 (<0.001)	−30.9 (<0.001)		5.46 (>0.999)	−65.69 (<0.001)	−9.89 (<0.001)
5	−34.83 (<0.001)	−66.52 (<0.001)	−34.53 (<0.001)	−5.46 (<0.001)		−69.43 (<0.001)	−11.85 (<0.001)
6	−0.47 (0.32)	−1.71 (0.044)	50.4 (>0.999)	65.69 (>0.999)	69.43 (>0.999)		35.82 (>0.999)
7	−24.39 (<0.001)	−36.46 (<0.001)	−8.48 (<0.001)	9.89 (>0.999)	11.85 (>0.999)	−35.82 (<0.001)	

The main number is the mean of the ES among the 731 rolling window for each model and each horizon. The number in brackets shows the corresponding standard deviation.
The Diebold Mariano test for the Energy score and the first horizon. The main number is the t-statistic value and in brackets the corresponding p-value.

Table 5. Energy score values and DM-test

Nonetheless, another reason could be the structural breaks in the data set lead to the poor forecasting performance. In the Table 5 it is observed that the models with bivariate jumps provide generally better forecasts with this score which means that the dependency structures are properly included by considering bivariate jump. This is a interesting results as our objective including bivariate jump is too improve the usual OrnsteinUhlenbeck process.

As it can be observed in Table 5 with ES it is not possible to distinguish between the OLS and enet estimation techniques forecasts while with the other scores the elastic net method is significantly better. We observe that the ARX-BiJ-μ_d model provides significantly better forecasts than the other models followed by ARX-BiJ. As already noted, according to the ES criterion the models with bivariate jump offer significantly better results which means that our adjustment on the jump diffusion models helps to capture the dependencies efficiently. It is clear that according to this score the assumption of no constant mean of the jump helps in the forecasting accuracy. This improvement was not that clear with the other scores. The forecasting performance of the ARX-GARCH model is weaker than expected, the reason could be that CCC-GARCH structures are more focused on symmetric effects.

7 Conclusion

The proper modelling and forecasting in electricity markets is crucial for all the participants. In this paper we focus on the off-peak and peak time series. These time series are the ones traded in the future markets, hence it is really important to have accurate forecasts in order to avoid losses.

Our motivation in this paper is to include dependency structure to the bivariate analysis. We believe that it is extremely relevant that the forecasts preserve the correlation structure from the original time series. In the literature, so far, only have been applied MRJD to independent time series. Our approach is to include bivariate jump occurrences to the MRJD model. Afterwards, we need to evaluate whether these correlation structures have been properly included or not, in order to do so we need a criterion which takes into account the dependencies. In our case we use the ES to that purpose but it is not the only criterion we apply, we use the MAE, MSE and the PB, as well. Additionally, we apply the DM test to compare the models in pairs for each one of the horizons.

Regarding the results we observe that the models with bivariate jump are not forecasting better according to the MAE and MSE criteria. However, with the PB score criterion where the distribution of the forecasts is evaluated the performance is slightly better when bivariate jump structure is considered in the model. When we focus on the ES score we observe the the models with correlated jumps perform significantly better than the remaining models. Nonetheless, the most complex model which deals with bivariate jumps, no constant jump size and CCC-GARCH structure does not outperform the forecasts of the same model but without considering the CCC-GARCH. We assume this happens because there are too many parameters to estimate in the second step and the optimal value is a local maximum or the presence of structural breaks in the data set which lead to the poor performance.

For further research it might interesting to develop dependency structure models considering hourly data and make a multivariate analysis with 24 variables. The problem with this 24 variables is that the number of parameters to estimate increases too much.

Bibliography

[1] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International journal of forecasting*, 30(4):1030–1081, 2014.

[2] Florian Ziel. Forecasting electricity spot prices using lasso: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems*, 31(6):4977–4987, 2016.

[3] Nektaria V Karakatsani and Derek W Bunn. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24(4):764–785, 2008.

[4] Florian Ziel, Rick Steinert, and Sven Husmann. Efficient modeling and forecasting of electricity spot prices. *Energy Economics*, 47:98–111, 2015.

[5] Bartosz Uniejewski, Jakub Nowotarski, and Rafał Weron. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, 9(8):621, 2016.

[6] Florian Ziel and Rafał Weron. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, 70:396–420, 2018.

[7] Dogan Keles, Massimo Genoese, Dominik Möst, and Wolf Fichtner. Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. *Energy Economics*, 34(4):1012–1032, 2012.

[8] George E Uhlenbeck and Leonard S Ornstein. On the theory of the brownian motion. *Physical review*, 36(5):823, 1930.

[9] Rafał Weron. Market price of risk implied by asian-style electricity options and futures. *Energy Economics*, 30(3):1098–1115, 2008.

[10] Alvaro Cartea and Marcelo G Figueroa. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005.

[11] Bin Dai, Shilin Ding, Grace Wahba, et al. Multivariate bernoulli distribution. *Bernoulli*, 19(4):1465–1483, 2013.

[12] Sergey Voronin, Jarmo Partanen, and Tuomo Kauranne. A hybrid electricity price forecasting model for the nordic electricity spot market. *International Transactions on Electrical Energy Systems*, 24(5):736–760, 2014.

[13] Katarzyna Maciejowska and Jakub Nowotarski. A hybrid model for gefcom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, 32(3):1051–1056, 2016.

[14] Grzegorz Dudek. Multilayer perceptron for gefcom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, 32(3):1057–1060, 2016.

[15] Romain Juban, Henrik Ohlsson, Mehdi Maasoumy, Louis Poirier, and J Zico Kolter. A multiple quantile regression approach to the wind, solar, and price tracks of gefcom2014. *International Journal of Forecasting*, 32(3):1094–1102, 2016.

[16] Pierre Pinson and Robin Girard. Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, 96:12–20, 2012.

[17] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[18] Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and statistics*, pages 498–505, 1990.

[19] Annastiina Silvennoinen and Timo Teräsvirta. Multivariate garch models. In *Handbook of financial time series*, pages 201–229. Springer, 2009.

[20] Giovanna Zanotti, Giampaolo Gabbi, and Manuela Geranio. Hedging with futures: Efficacy of garch correlation models to european electricity markets. *Journal of International Financial Markets, Institutions and Money*, 20(2):135–148, 2010.

[21] Helen Higgs. Modelling price and volatility inter-relationships in the australian wholesale spot electricity markets. *Energy Economics*, 31(5):748–756, 2009.

[22] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[23] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[24] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[25] Pierre Gaillard, Yannig Goude, and Raphaël Nedellec. Additive models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting. *International Journal of forecasting*, 32(3):1038–1050, 2016.

[26] Florian Ziel and Rick Steinert. Electricity price forecasting using sale and purchase curves: The x-model. *Energy Economics*, 59:435–454, 2016.

[27] Ingo Steinwart, Andreas Christmann, et al. Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17(1):211–225, 2011.

[28] Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 2002.

[29] Francis X Diebold. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–1, 2015.

Computing Environment for Forecasting based on System Dynamics Models

Radosław Pytlak[1], Damian Suski[2], Tomasz Tarnawski[3], Zbigniew Wawrzyniak[4], Tomasz Zawadzki[5], and Paweł Cichosz[4]

[1] Faculty of Mathemetics and Information Science, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland
r.pytlak@mini.pw.edu.pl
[2] Faculty of Mechatronics, Warsaw University of Technology
[3] Kozminski University, 03-301 Warsaw, Poland
[4] Faculty of Electronics and Information Technology, Warsaw University of Technology
[5] War Studies University, 00-910 Warsaw, Poland

Abstract. The paper proposes the computing environment which allows building forecasts based on System Dynamics models. The environment is equipped with GUI for building dynamic models according to the Forrester's methodology. The unique feature of the application is that the created models can be calibrated with the help of optimization procedures tailored for solving nonlinear least squares problems with differential–algebraic equations. Furthermore, the application enables verification of decision rules inherited in System Dynamics models by solving associated with the models dynamic optimization problems—then the model with optimal decision rules can be simulated to build forecasts of interests. To illustrate the functionalities of the environment the example of the model of drug prevalence is discussed in some detail.

Keywords: System Dynamics, trajectories forecasting, decision rules, optimal control

1 Introduction

System Dynamics offers an approach to forecasting that is essentially different from the one based on econometric models, as it allows for including decision rules (which are dynamical in nature) within the model. Building predictions boils down to running new simulations of model equations with appropriately changed parameters describing the decision rules. Also, nonlinear functions are a common thing in multiequation System Dynamics models while multiequation nonlinear econometric models are a rarity due to extreme difficulties in their building.

The quality of quantitative predictions drawn form System Dynamics models – in the form of generated trajectories of selected variables – depends heavily on the specification of values of the model's parameters. If our predictions are to

possess any merit we must be certain that the simulated trajectories of interest are close to these which have been observed. This can be achieved in the process of model calibration, provided that the structure of model's equations are not far from the actual dynamics of the real process. Model calibration can refer to statistical methods to measure quality of parameters estimation, but the final model and its use in predicting future values of model's variables is deterministic. The model calibration is based on optimization tools, however there are other applications of optimization within the process of building dynamical models, such as finding adequate structures of model's decision rules or choosing proper (parameters') values for these decision rules.

The paper describes a newly implemented computing environment that brings together System Dynamics (SD) graphical modeling methodology, simulation of the model's behavior, defining optimal control problems (by supplementing the SD model with objective function) and advanced numerical algorithms for solving various classes of dynamic optimization problems. It was designed to make it possible to define and solve a number of related problems defined around the same model. This way one can start with specifying a dedicated problem of calibrating the model's parameters (solved by selecting parameters' values for which the distance between empirical and generated trajectories is minimized). Once the model is calibrated, it can then be used to build another optimization problem dealing with choosing control strategies that would optimize (some measures of) certain aspects of the system's behavior.

The proposed software package can be compared to a number of existing solutions, such as Vensim ([16] or any of similar SD simulation applications), OpenModelica [15] or JModelica.org [14]. However, we believe that our solution is unique as it exceeds the capabilities of these competitors. In comparison with Vensim – it not only offers simulation (and does that with a significantly wider selection of numerical integration engines) but also provides support for dynamic optimization, which is completely lacking in the Vensim-class of System Dynamics simulation packages. In that respect it is much closer in kind to the Modelica-related applications with built-in Optimica language support, i.e. OpenModelica or JModelica.org (Optimica is a close relative to Modelica with slight syntax extensions for defining optimal control problems; it was initially proposed in [1]). What sets us apart from the two, is the designed-in support for a wide range of various optimization solvers (as described e.g. in [9]) – while both implementations, of OpenModelica and JModelca.org, are based solely on the collocation algorithm – which makes them less appropriate for certain classes of problems.

In our environment the models can either be built by means of System Dynamics standard graphical notation or through text files written in a dedicated language – DOML. The Dynamic Optimization Modeling Language (DOML) was initially proposed in [9] as a Modelica-based programming-language-independent communication format while developing Interactive Dynamic Optimization Server (IDOS, described therein). The DOML format provides a mean for defining dynamic systems (i.e. described by systems of Ordinary Differential

Equations or Differential-Algebraic Equations) together with user controls and objective functions that together compose optimal control problems.

The paper is organized as follows. The following section 2 provides a general description of the graphical and computational environment. Section 3 proceeds towards formulating an exemplary dynamical model that illustrates its capacities and can be used for forecasting the behavior of a system of interest. The example model deals with forecasting drug use, at a national level, under different strategies undertaken by law enforcement authorities to combat drugs. Next, in section 4, the numerical algorithms implemented within the environment and used in this particular case are briefly described, together with discussion of the related aspects of calibrating the model and specifying decision rules. Finally, section 5 provides a discussion of simulation and optimization results yielding trajectories showing the foreseen behavior of the system's variables.

2 System Description

The system that has been developed to solve optimization problems consists of three main components:

- modeling application
- DOML compiler
- simulation and optimization libraries.

and their inter-relationship is schematically shown on Figure 1.

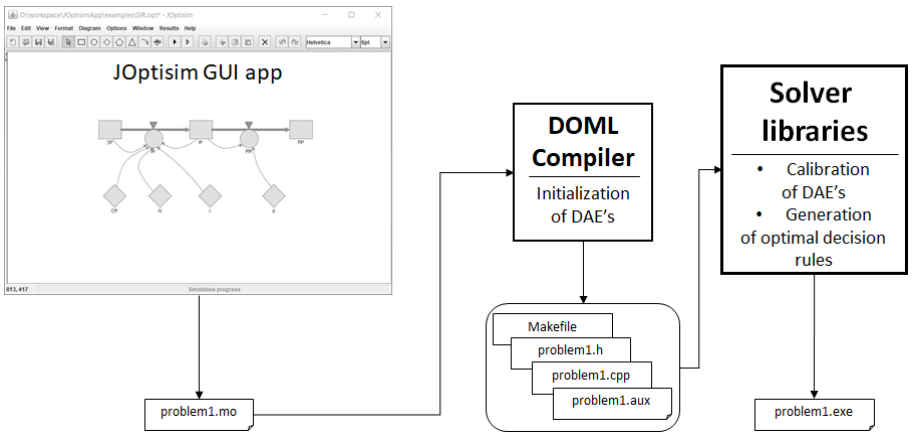


Figure 1: Composing elements of the computing environment

The modeling component called JOptisim allows users to create graphical models according to System Dynamics methodology which introduces three fundamental elements of SD models, namely stocks, flows and auxiliary variables (or

parameters) – in that respect it is comparable to similar graphical tools such as Vensim, Stella, etc as all model's elements can be graphically placed, removed, renamed or copied at any time and then saved to and restored from external file. JOptsim also runs preliminary simulations to check correctness of the model. The implemented simulation procedure is based on advanced RADAU5 numerical code which enable to track trajectories with steep changes, especially observed in so called stiff equations.

Besides creating a standard System Dynamics models, application allows to define optimization problem by filling additional values in a dedicated window. The window supports a definition of: horizon used in optimization problem; objective function; constraints of optimization problem. Such created model can be saved in an external file for further editing or exported to the DOML language file either as ordinary differential or differential-algebraic equations. The resulting DOML files are sent to Modelica-based simulation and optimization engine composed of DOML compiler (as the first stage) and a number of dedicated numerical libraries.

The compiler was fitted with a number of additional features needed to implement the postulated extensions to the standard Modelica language. The extensions and their implementations were discussed in detail in a number of articles (see e.g. [8], [9]) but in short they boil down to:

- providing a mechanism for choosing among a number of different solver packages applicable to (different classes of) dynamic optimization problems;
- allowing for solver chaining – i.e. using a sequence of solvers on the same problem, where the solution from one (usually more robust but less accurate) can be used in the next (that e.g. provides higher accuracy but requires a relatively good starting point);
- introducing labeling of equations (and constraints) together with a way for defining so called adjoint variables, which are necessary when using certain class of optimization algorithms.

To our mind, these elements make the language significantly better fit for handling the vast variety of optimal control problems out there.

The compiler transcribes the DOML definition of a problem into a regular programming language (in most cases being C++, but in some cases it is C, Fortran of R). The generated code includes appropriately woven calls to external numerical libraries needed to solve a particular problem. The environment set-up is prepared to handle various cases as it is equipped with several sophisticated solvers for dynamic optimization problems: based on apriori discretization of differential equations; solvers using adjoint equations; based on shooting procedures. Secondly, the environment includes state-of-the-art numerical libraries dealing e.g. with numerical integration or automatic differentiation.

3 Problem Formulation

The exemplary problem we consider is determination of the optimal time distribution of police forces intended for fighting against drugs. The calculations

are based on the drug prevalence model built with the help of System Dynamics methodology. The model is based on the drug prevalence model presented in [4] and additional contribution related to drug market in Poland depicted in [5].

The casual loop diagram of the appropriate model is presented in Fig. 2 and

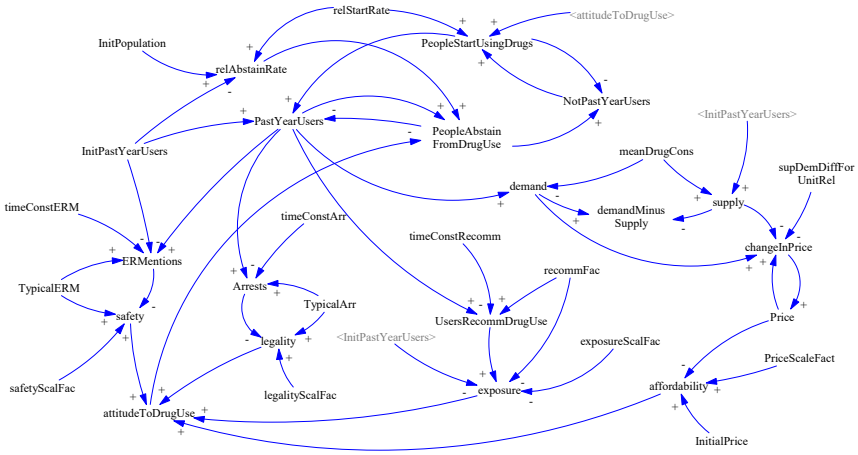


Figure 2: Casual loop diagram of System Dynamics model of drug prevalence

we will now briefly describe the structure of the model. The core of the model is the flow of people between two groups of people: people who have used drugs during the previous year (*PastYearUsers*); people who have not used drugs during the previous year (*NotPastYearUsers*). The flow of people between these two groups consists of two subflows: flow of people who have started using drugs during the previous year (*peopleStartUsingDrugs*); flow of people who have abstained from drug using during the previous year (*peopleAbstainFromDrugUse*). The sizes of these flows depend on the sizes of populations and the positive attitude to drug use among the society, which is modeled by one aggregated variable *attitudeToDrugUse*. In our model the attitude to drug use depends on the following factors: exposure to drugs through social contacts (*exposure*); perception of health consequences of drug use (*safety*); perception of consequences of breaking the drug law(*legality*); affordability of drugs (*affordability*).

The exposure to drugs through the social contacts depends on the prevalence of drugs in population and it increases the positive attitude to drug use. In our model we assume that the perception of health consequences of drug use depends

on the number of medical interventions associated with using drugs (*ERMentions*). With the increasing number of medical interventions the positive attitude to drug use decreases. The number of medical interventions depends directly on the prevalence of drugs in population. The perception of law consequences of drug possession and distribution depends on the number of arrests (*Arrests*). With the increasing number of arrests the positive attitude to drug use decreases. The number of arrests depends on the prevalence of drugs in population as well as on the efficiency of police (*policeEfficiency*). The police efficiency models a general ability of police to fight against drugs and it is assumed to be the only control variable in our exemplary problem.

Drugs affordability depends on the drug unit price (*Price*). With the increasing price the positive attitude to drug use decreases. To describe the dependence of drug unit price on the drug prevalence a simple economic modeled is used. It is assumed that the changes of drug unit price depend on the actual price and the difference between drug demand and supply. Drug demand depends on the drug prevalence and drug supply is modeled as an external function of time. It is possible to utilize more complicated economic models, but for our needs the proposed model gives satisfactory results.

Our goal is to determine the optimal distribution of police forces over the assumed time interval to achieve the best results in fighting against drugs. In our model the optimized control variable is *policeEfficiency*, which represents the efficiency of the police in arresting people breaking the drug law. *policeEfficiency* variable has been introduced to our model arbitrary and does not possess a simple and direct interpretation. Nevertheless it is possible to associate it with other measurable quantities such as a number of police patrols per week or the drug fight budget per year. In any case the methodology remains the same.

To test the computational environment we solved the exemplary optimal control problem. The parameters of the drug prevalence model have been adjusted according to data regarding the cocaine prevalence in Poland provided by European Monitoring Centre for Drugs and Drug Addictions. In the considered optimal control problem we assume that at each time moment the *policeEfficiency* stays within interval $[0, 2]$. This constraint models e.g. the limited number of policemen available at each moment. What is more we assume that the following inequality holds $\int_0^{20} \mathit{policeEfficiency}(t)dt \leq 20$. The considered time interval is $[0, 20]$ years. The above inequality represents the budget constraints as the *policeEfficiency* cannot always assume the maximum value, for which the integral would be 40. It is therefore required to distribute the police forces somehow over the whole time interval. The proposed cost function, which is supposed to be minimized, is the number of drug users at the final time $\mathit{PastYearUsers}(20)$.

The model and the optimization tasks have been implemented as a DOML script [9]. The header of the DOML script defines the cost function and the time interval.

```
optimization drugs_opt ( objective = PastYearUsers(finalTime),
startTime = 0, finalTime = 20.0 )
```

Next all the variables in the model are declared.

```
parameter Real InitialPrice= 45; ...
Real affordability; ...
input Real policeEfficiency(initialGuess = 1, min = 0, max = 2);
...
```

There are four type of variables: constant parameters, casual variables, input variables, and state variables, which correspond to stock variables in System Dynamics convention [13]. Besides the declaration, for input variables the initial guess and bounds are defined and for state variables the initial value is defined.

In the `equation` section of DOML script system equations are defined. Differential equations defines derivatives of state variables, whereas algebraic equations defines the remaining variables.

```
equation
  der(Price) = (changeInPrice) - 0; ...
  0 = - changeInPrice + (demand-supply)/supDemDiffForUnitRelChange
      * Price;
```

In the `constraint` section the remaining constraints can be defined. The integral inequality is defined by introducing the additional variable *CostOfPoliceEffciency* which integrates *policeEfficiency* over time. Then, the constraint section contains

```
constraint
  CostOfPoliceEffciency(finalTime) <= 20;
```

4 Numerical procedures for optimal control problems

There are three types of optimization problems linked with the methodology our computing environment realizes. The first two are associated with models calibration. We assume that the calibration of a System Dynamics model should be performed in two stages. In the first one one look for functions which describe the relationship in a given node of a CLD diagram. This calibration reduces in fact to solving a standard nonlinear (in general) least squares problem and can be accomplished by a variant of the Gauss–Newton method. In that case we use data which represent variables linked to a particular node.

The model calibration in the second stage is needed to construct an adequate model for forecasting with respect to a particular variable which is a solution to DAEs representing System Dynamics model. In that case the calibration procedure is much more elaborate since it takes into account the entire model dynamics and so it requires a tailored dynamic optimization procedure based on a least squares objective function.

Then, we apply optimal control to verify (construct) decision rules which are inherently present in a System Dynamics model. In that case we solve a standard optimal control problem having in mind that a procedure for solving that problem must satisfy the following conditions: 1) it should tackle control problems described by large–scale differential–algebraic equations; 2) it must

have an initialization procedure for algebraic variables which is called at initial time and at times at which controls exhibit jumps.

The first requirement is a direct consequence of the System Dynamics methodology. The dynamic model is the form of diifferential–algebraic equations of the special form. Alebraic variables are functions of other algebraic variables and differential variables (those variables which are solutions to differential equations). In the literature on differential–algebraic equations of this form are called semi–explicit index one differential–algebraic equations ([2]).

The solver we use to solve optimal control problems with System Dynamics models is described in detail in [7]. The solver has the following features:

(a) it is based on an implicit Runge–Kutta procedure with variable stepsizes–it implies that it can tackle *stiff* differential–algebraic equations;
(b) the Runge–Kutta integrator requires the description of system equations jacobians as sparse matrices thus it is suitable for integrating large–scale differential–algebraic equations;
(c) the solver is based on a continuous time model of system equations which means that it is appropriate to problems with state constraints which can exhibit both *touch points* and *active arcs* at solutions—parameters of these active sections of state constraints can only be precisely determined if the continuous time model is applied;
(d) it is an SQP type algorithm which assures fast convergence but at the same time does not refer to Lagrange multiplires associated with constraints–the Lagrange multiplires corresponding to state constraints may be difficult to evaluate.

4.1 Nonlinear least squares calibration of dynamical systems

In order to explain our approach to the calibration of dynamical models consider first the static nonlinear least squares problem:

$$\min_{p\in\mathcal{R}^{n_p}} \left[f(x) = \frac{1}{2}\|g(p)\|^2 = \frac{1}{2}\sum_{i=1}^{m} \left((g_i(p)\right)^2 \right] \tag{1}$$

where $g(p) = [g_1(p), g_2(p), \ldots, g_m(p)]^T$.

Important feature of the Gauss–Newton method for the nonlinear least squares problem is that the Hessian matrix of the objective function (1) can be effectively approximated by the matrix based on the Jacobian matrix related to the transformation g—see [6] for details:

$$\nabla^2 f(p) \approx J(p)^T J(p). \tag{2}$$

Here, $J(p)$ is the Jacobian matrix of the transformation g.

Let us adopt the Gauss–Newton procedure for the one suitable for calibrating dynamical systems with continuous time dynamics described by DAEs.

Suppose that a system is represented by the equations

$$F(\dot{x}(t), x(t), u(t), p) = 0, \quad t \in [0, t_f], \tag{3}$$

where vector p represents system parameters. The solution to system (3) is dependent on p and thus we denote it by x^p.

We approximate the parameters p by solving the nonlinear least squares problem

$$\min_p \left[MSE_l(p) = \frac{1}{N} \sum_{j=1}^{N} \left(x_l^p(t_j) - \hat{x}_l(t_j) \right)^2 = \frac{1}{N} \sum_{j=1}^{N} MSE_l^j(p) \right] \tag{4}$$

subject to the constraints (3). Here, we recall, that $\hat{x}_l$ is the empirical (measured) trajectory.

That problem can be solved by nonlinear programming techniques provided that we can evaluate gradients of $MSE_l(p)$.

On the other hand the gradient of the functional $MAE_l(p)$ may be evaluated with the help of adjoint equations if we observe that $MSE_l(p)$ is composed of N functions $MSE_l^j(p)$.

The gradient of $MSE_l^j(p)$ can be calculated by using the adjoint equations for the system (3) and the function $MSE_l^j(p)$—it means that these adjoint equations will be integrated from time t_j (see, for example, [7] on the use of adjoint equations in dynamic optimization and [8] on their application to this particular function). Then, $\nabla MSE_l(p) = \sum_{j=1}^{N} \nabla MSE_l^j(p)$. Having objective function values and its gradients we can build an optimization procedure for model calibration. A general scheme for such a procedure may look like stated below.

General Calibration Procedure (GCP)

1. Set initial values of parameters: p_1 and set $k = 1$.
2. For parameters p_k calculate system trajectories x^{p_k} by numerically integrating system equations. On that basis determine objective function value $MSE_l(p_k)$ through values $MSE_l^j(p_k)$, $j = 1, \ldots, N$.
3. Having trajectories x^{p_k} and values $MSE_l^j(p_k)$, $j = 1, \ldots, N$ solve adjoint equations, and determine $\nabla MSE_l^j(p_k)$, $j = 1, \ldots, N$, $\nabla MSE_l(p_k)$.
4. Determine the direction of descent p_k using some optimization procedure.
5. Perform directional minimization with the help of the optimization procedure to evaluate the step size α_k. Substitute $p_k + \alpha_k d_k$ for p_{k+1}. increase k by one and go to Step 2).

That general calibration scheme covers also a Gauss–Newton approach to a model calibration. However, in this case we need specyfing how the direction of descent d_k is determined on the basis of vectors $\nabla MSE_l^j(p_k)$, $j = 1, \ldots, N$ and

$\nabla MSE_l(p_k)$. In this case we use the relations:

$$J_l(p_k) = \begin{bmatrix} \nabla MSE_l^1(p_k)^T \\ \nabla MSE_l^2(p_k)^T \\ \vdots \\ \nabla MSE_l^N(p_k)^T \end{bmatrix}, \; H_l(p_k) = J_l(p_k)^T J_l(p_k)$$

$$p_{k+1} = p_k - \alpha_k \left[H_l(p_k)\right]^{-1} \nabla MSE_l(p_k). \tag{5}$$

4.2 Algorithm for optimal control problems with SD models

In general System Dynamics models are described by differential–algebraic equations (DAEs). Therefore, we should consider the following optimal control problem $\mathbf{P_{DAE}}$:

$$\min_u \phi(x(t_f)),$$

subject to the constraints:

$$F(\dot{x}(t), x(t), y(t), u(t), t) = 0 \text{ a.e. on } T, \; x(0) = x_0 \tag{6}$$

$$q(t, x(t)) \leq 0 \;\; \forall t \in T \tag{7}$$

$$h_i^1(x(t_f)) = 0 \;\; \forall i \in E \tag{8}$$

$$h_j^2(x(t_f)) \leq 0 \;\; \forall j \in I \tag{9}$$

$$u \in \mathcal{U} = \{u : \; u(t) \in \Omega \text{ a.e. on } T\}. \tag{10}$$

Here, $x(t) \in \mathcal{R}^{n_d}$, $y(t) \in \mathcal{R}^{n_a}$, $u(t) \in \mathcal{R}^m$, $n = n_d + n_a$ and Ω is a convex compact set. We assume that for any x_0, $u \in \mathcal{U}$ there exists a unique solution to (6): (x^u, y^u). We call x a *differential* state, and y an *algebraic* state.

The problem we have to consider has both terminal constraints (8)–(9) (equality and inequality), state constraints (7) (constraints which must be satisfied at every point of time horizon) and hard constraints on controls (10). The system dynamics equations (6) contains controls u which are introduced to System Dynamics to verify optimality of decision rules used in the System Dynamics model. Once the problem $\mathbf{P_{DAE}}$ is solved we build once again (possibly new, but almost surely with new parameters) decision rules and new forecasts on the basis of these rules.

Essentially the above optimal control problem can be numerically tackled in the very similar way as the problem described by ODEs, in particular, due to the uniqueness assumption stated above, it can be stated in *reduced space* (see [7] for details). In fact our computing environment applies a numerical procedures stated in [7]. It is the procedure which uses a RADAU IIA procedure (it is an implicit Runge–Kutta method) for the integration of the system equations (6)— the justification for the choice is given in [7] (Chapter 6). Having trajections of system equations obtained by numerical integration gradients of all functionals

defining the problem $\mathbf{P_{DAE}}$ are evaluated in the reduced space with the help of adjoint equations associated with system equations (Chapter 6 of [7]).

Since our system equations are in the form of DAEs we need a procedure for their initialization—usually we have initial values of differential states x but algebraic variables y have to be found so algebraic equations are satisfied at initial time. Due to the way System Dynamics equations are constructed (with the help of the GUI) the functions $y_i(t) = init(x^u(t), u(t), i)$, $i = 1, \ldots, n_a$ can be built—these functions are created while processing model equations by the DOML compiler.

5 Discussion of Results

Several optimal control problems based on System Dynamics have been defined and solved. In [10] an optimal control problem related to sanitary teams activities during an epidemics of a foodborne disease is discussed. In the problem decision variables (controls) are linked with the number of sanitary teams which try to contain the epidemics by isolating infected people and by eliminating contaminated food. The optimal control problem is further discussed in [12] where decision rules are derived for sanitary teams activities on the basis of the optimal solution to the problem.

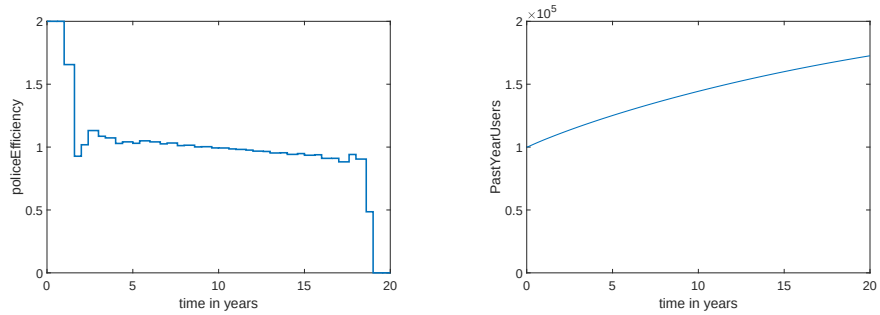
The optimization results presented in the section are related to the System Dynamics model of drug prevalence stated in Section 3. They should be treated as preliminary results since additional experiments with the optimization model (and its variants) are needed to propose new decision rules for the distribution of police forces which fight against drugs use. On the basis of these decision rules we could then forecast the spread of drugs use as described by the model variable *PastYearUsers*.

The numerical procedure presented in Section 4 has found the optimal solution shown in Fig. 3a. The analysis of the obtained solution suggests the following strategy for the use of police forces: try to apply police forces at its maximum efficiency in the beginning (for a few years); then stick to the slightly decreasing average efficiency until resources are exhausted as expressed by the constraint on the final value of *CostOfPoliceEffciency* —parameters of this strategy could be obtained by solving another optimization problem in which its parameters are decision variables. If that strategy is used then the number of dugs users will be minimized at the end of considered horizon.

Acknowledgments. The work was partially funded by the grant DOB-BIO7/05 /02/2015 of Polish National Office for Research and Development.

References

1. Åkesson J. Tools and Languages for Optimization of Large-Scale Systems. Lund, Sweden: *PhD thesis*, Department of Automatic Control, Lund University, 2007.



(a) Optimal trajectory for the control *policeEfficiency*

(b) Optimal trajectory for the variable *PastYearUsers*

Figure 3: Optimal trajectories.

2. Brenan, K.E., Campbell, S.L., Petzold, L.R.: Numerical Solution of Initial-Value Problems in Differential–Algebraic Equations. SIAM (1996).
3. Bryson, A.E., Ho, Y.: Applied Optimal Control: Optimization, Estimation and Control, Taylor & Francis (1975).
4. Homer, J.B.: A system dynamics model of national cocaine prevalence. Sys. Dyn. Rev. 9(1), 49-78 (1993).
5. Hołyst, B., Wawrzyniak, Z. M., Wiciak, K. (2017). Prognozowanie kryminologiczne w wymiarze społecznym, Tom I–Metodologia, Analiza, Tendencje rozwojowe (Vol. 1). Wydawnictwo Naukowe PWN SA (in Polish).
6. Nocedal, J., Wright S.: Numerical Optimization. Springer, New York (1999).
7. Pytlak, R.: Numerical Methods for Optimal Control Problem With State Constraints, Lecture Notes in Mathematics, 1707, Springer–Verlag, Berlin, Heidelberg (1999).
8. Pytlak, R., Tarnawski, T.: IDOS - (also) a Web Based Tool for Calibrating Modelica Models. In: Proceedings of the 10th International Modelica Conference, March 10-12, Lund, Sweden, pp. 1095–1104 (2014).
9. Pytlak, R., Tarnawski, T., Fajdek, B., Stachura, M.: Interactive dynamic optimization server–connecting one modeling language with many solvers. Optimization Methods & Software 29, 118–1138 (2014).
10. Pytlak, R., Suski, D., Zawadzki, T., Stecz, W.: Decision Support System for Sanitary Teams Activities, International Journal of Decision Support System Technology 3, 23–45 (2014).
11. Pytlak, R., Zawadzki, T.: On solving optimal control problems with higher index DAEs, Optimization Methods & Software 29, 1139–1162 (2014).
12. Pytlak, R., Zawadzki, T.: A New Approach to Optimization of a System Dynamics Model. submitted for publication.
13. Sterman, J.D.: Business dynamics, systems thinking and modeling for a complex world. The McGraw-Hill Companies, Inc, (2000)
14. Open Source Project: JModelica.org. `http://www.jmodelica.org/`. April 2017.
15. OpenModelica site, Open Source Modelica Consortium (OSMC). `http://www.openmodelica.org/`. June 2017.
16. Vensim site, Ventana Systems. `http://vensim.com/`. June 2018.

The Contrast Between Management Consulting and Outsourcing Management Services
A financial perspective

Leandro Pereira[1]; Carlos Jerónimo[1]; José Santos[2 ∞]; Nelson António[1]

[1] Business Research Unit – BRU-IUL, ISCTE, Portugal
[2] Winning LAB, Winning Scientific Management, Portugal
[∞] jose.santos@winning.pt

Abstract— In a world in which the most effective way to deliver services is often discussed and many times it is the customer who ends up defining the delivery model of the companies, it becomes important to analyze and compare the tendency and financial sustainability among the companies that deliver management services in a different format, whether that is through consulting or outsourcing services. This study was conducted by analyzing financial indicators such as EBITDA Margin between 2014 and 2017 and by analyzing the forecast between 2018 until 2023 from the top three consulting and outsourcing companies. The findings suggested that most organizations from Management Consulting Services show better results when we are referring to "EBITDA". Nevertheless, there are aspects that require additional analysis.

Keywords — Management consulting, Outsourcing, Performance measurement, profitability, decision

1 Introduction

In an era in which the speed of adaptation and even anticipation of the market cause bankruptcy of companies whenever they cannot meet their real needs, the lines of management are undergoing mandatory changes. Through the analysis of the gains of many outsourcing firms, there are many non-core activities undertaken by these companies that promise to their customers reducing costs and focus on their business activities that create direct value [1]. Much research has been done on how companies choose which activities to do internally or externally. There is a lot of repetition of simple cost/benefit arguments that are difficult to measure. What is the long-term cost of letting the company managers learn about how to make a strategic plan rather than hiring a specialized, referenced consulting firm on the subject? What is the benefit of using Outsourcing in 80% of our non-core activities? Do we need a company to help us think or to execute tasks? Or even both? Issues are not impossible, but they are difficult to justify in a standardized and scientific way for many organizations. On the contrary, for service providers the big question remains on knowing the best model to follow because if a certain model is so good, can it be improved to bring best turnovers? It is also true that there are some outsourcing companies that often only provide body shopping services and their business model is easily replicated and they have almost no competitive advantage [2]. This article has a main objective to carry out a financial performance analysis between management consulting services, management outsourcing and mixed model firms that until now have been somehow neglected by the academic world. For this, it was necessary to start with a study on the definition of the concept of management consulting and outsourcing, concepts that are constantly changing and have never been fully defined [3]. At the management consulting level, the concept of the market itself is not always defined in the same way and there is no specific skill totally defined, and even ironized by [3], who refers to it as "The method can be, and is, applied not only by full-time consultants, but also by many other technically competent persons whose main occupation may be teaching, training, research, systems development, project development and evaluation, technical assistance to developing countries, and so on". According to a report released in 2010 [4] by the European Federation of Management Consulting Associations (FEACO), which represents 23 national associations across Europe including UK's Management Consultancies Association has been proposed in 2005 a new definition of the scope of services offered by management consulting firms in Europe and a new way of segmenting these services as following: (1) Business Consulting and (1b) IT Consulting, (2) Development and Systems Integration (excluding software development), (3) Outsourcing of value added services and (4) other services. According to Fig. 1 Business Consulting is the most robust service line with 50% of Management Consulting. The Outsourcing market represents 12%.

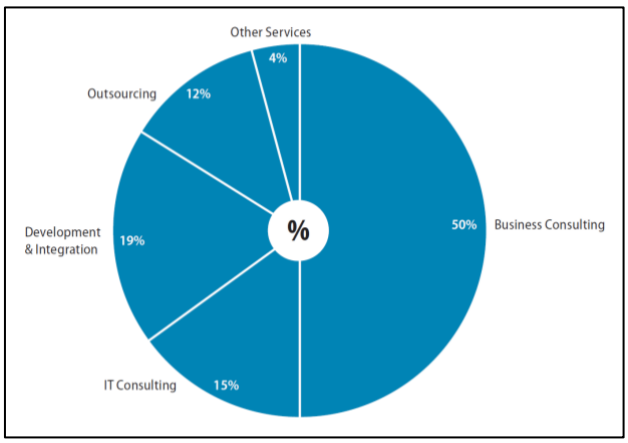


Fig. 1 - Breakdown of turnover by service line (Source: FEACO, 2010)

After defining and reviewing the services of companies in a scientific way and because it was not possible to separate them, a survey was carried out by a group of experts to guarantee greater definitions of this services. After the typification, it was carried out the financial analysis based on the country, CAE and name. The findings revealed, according to the data provided, that consulting firms have a better performance when compared with companies with mixed services or unique outsourcing services.

1.2 Academic Perspective

Management Consulting

The Management Consulting concept is still relatively young, with just over 100 years old. The issues of counseling and problem solving have been known since the end of the industrial era [3] and subjects such as lack of experience in organizing processes, people or demands of mass-produced goods have led some experts like Frederick Winslow Taylor and Charles Babbage to develop new methods for work organization, which resulted in significant improvements [5]. Arthur D. Little, was a professor that created in 1886 the first recognized management consulting firm even though he concentrated only on more technical terms. Later, around 1900, Arthur Young, William Deloitte and George Touche, advanced with more themes of management and auditing. Around 1914, Booz Allen Hamilton formed the first management consulting firm to serve the industry as well as government clients while James O. McKinsey and Andrew Kearney founded McKinsey & Company in 1926 and became the first Management Consulting firm with the most modern strategy. Robert David [6] suggests that the growth that was experienced during the 1930s and 1980s was induced by four external forces, being them the increasing number and complexity of firms, the ideology of corporate conversion to non-corporate sectors, the organizational efforts of the second World War and finally the increasing impact of business education and business press.

The evolution of management consulting is a process that occurred in three specific periods, because of the changes in the environment where organizations operate, as well as due to the changes in management and the evolution of the organizations themselves, and as a way of accompanying the trends. Nonetheless, there are divergent opinions regarding the specific periods when those phases took place. According to some authors [7-8-9-10-11-12,13], the professional development of management consulting firms can be fragmented in scientific management, which appeared in the turn of the twentieth century and was centered on the efficiency of the company, also known as "shop-floor efficiency". It was followed by organization and strategy with companies providing advice on organizational issues, exemplified by Arthur D. Little, Booz Allen & Hamilton and McKinsey. It finally shifted to communication and information technology (since 1990), exemplified by the largest accounting and IT firms from the 1970s onwards [13].

According to the FEACO [4], the Management Consulting market is divided into four segments as shown in the figure 2 bellow.

Service Line	Sub Service	Description
1.Consulting	1a. Business Consulting (BC)	• **Strategy Consulting (SC)**, which targets the improvement of the long-term, strategic health of a company: strategic planning development; mergers & acquisitions; sales; marketing; corporate communication; financial advisory; HR strategy; • **Organisation/Operation Management (OM)** aims at the integration of business solutions through Business Process Re-engineering (BPR); customer/supplier relations management (CRM); turnaround/cost reduction and purchasing & supply management as well as advise on outsourcing; • **Project Management (PM)**, the application of knowledge, skills, tools and techniques to a broad range of activities in order to meet the requirements of a particular project; • **Change Management (CM)**, this consists of services which, on top of any other type of consulting service, help an organisation deal with the effects that change has on the human element of the organisation; • **Human Resources Consulting (HR)**, consulting services which target the improvement of the 'people' element of an organisation through performance measurement and management, reorganisation of benefits, compensations and retirement schemes, HR strategy and marketing, the development of talent strategies and executive coaching.
	1b. Information Technology Consulting (ITC)	Helps organisations to evaluate their IT strategies with the objective of aligning technology with the business process. These services include strategic planning and conceptions, operations and implementations.
2. Development and Integration		• **Development of applications** (excluding software); • **Creation of new functionalities** through, often tailored, process developments. Usually these developments integrate or unite internal or external business processes and can involve a conversion of applications so that they can be used for different platforms or conceptions, the development of applications (excluding software).
3.Outsourcing		• **IT management services;** among these are services for the operation of infrastructures (operation of systems, administration and security, follow up of cost-effectiveness, configuration management, management of technology, etc.) applications management, and help desk management; • **Applied Management Services (AMS);** this concerns the outsourcing of the development and implementation of support services for hardware, applications, CRM and infrastructures (tools for the development of applications and middleware, as well as software for information management, storage or systems and networks); • **Business Process Outsourcing (BPO);** this service supposes the externalisation of a complete business process.
4.Other Services	Consulting, Development and Integration, and Outsourcing	• **Training;** • **Engineering Consulting;** • **Outplacement, Executive Selection and Recruitment;** • **Audit and Accounting**.

Fig. 2. Service lines (Source: self-constructed table based on FEACO 2010 REPORT)

The increase of globalization promoted the boom in consulting, favoring a growth in the number of management consulting firms, and bringing a methodical approach to the study of strategy and management. It is important to mention that consultancy expanded due to the economic development and industrial era, which first began in the United States of America, got into Europe and finally got as well into the rest of the world. In the 90's there was another expansion of Management Consulting, firstly due to Central Europe opening its borders, and secondly, to the explosive growth of the World Wide Web. This industry suffered stagnation between 2001 and 2003 caused by the high-tech/e-business bubble crash that popped with severe consequences to clients and their consultants. The credit crunch (2009-2011) was considered another setback to consulting firms. Nowadays the consulting sector is a multi-billion-euro industry, which employs millions of people worldwide. As represented in Figure 3 and 4, during the period of 2013–2016 the European Management Consulting market experienced stable growth, as well as employment.

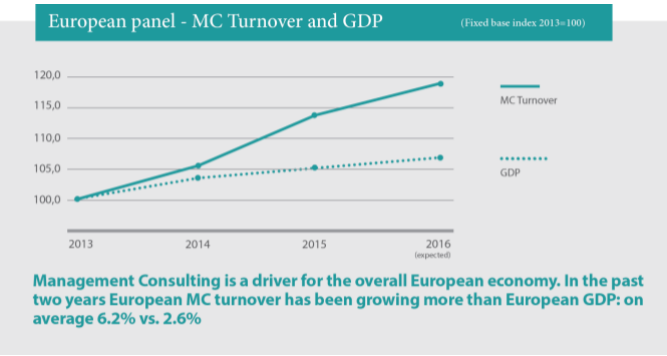


Fig. 3 - Turnover and GDP (Source: FEACO [14])

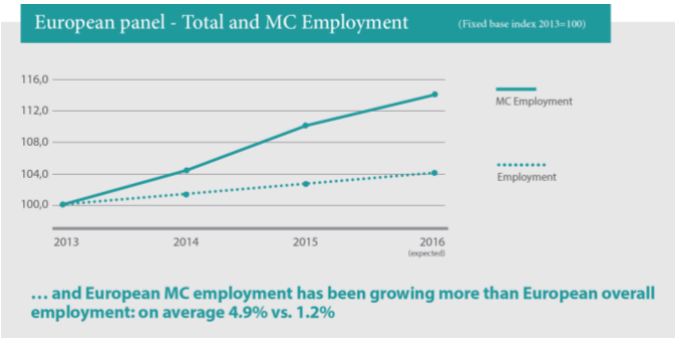


Fig. 4 - MC employment (Source: FEACO [14])

According to Fincham, Clark., Handley and Sturdy [9],[15], some companies started to believe that consultancy costs were too high when compared to the value generated, which means that the management consulting firms needed to evaluate and develop the value chain and be able to provide to customers methods, tools and above all tangible benefits of what they would earn, and explain that the value of things is measured not by cost but by perceived benefit [16].

Outsourcing – Complement or Integration

It has been mentioned and it is visible in **Error! Reference source not found.** that for FEACO outsourcing is an integral part of the management consulting component. However, there are other associations like FEACO and exclusively related to outsourcing such as the International Association of Outsourcing Professionals (https://www.iaop.org) and the Outsourcing Institute (http://outsourcing.com/). In fact, there are thousands of articles with the word outsourcing written in the title and without containing the word consulting (to date 78200 listed in Google Scholar). Probably this is also because the Management Consulting concept already mentioned above, also presents different concepts and it was possible to find in literature several generalist information's. Outsourcing and IT outsourcing, generally speaking, reflects "the use of external agents to perform one or more organizational activities" [17]. The author Monica Belcourt [18] states that Outsourcing happens when a company contracts another company to provide services or products of a noteworthy capacity or movement. That kind of work is customarily done inside and is moved to an outer supplier. Outsourcing contrasts from alliances or partnerships, here are profit sharing. Outsourcing is defined [19] as the act of obtaining semi-finished products, finished products or services from an outside company if these activities were traditionally performed internally. In the previous sentence, the word 'product' may be replaced by 'service'. The company that outsources is called 'buyer', whereas the company that provides the service is known as the 'vendor'. Some authors also show [20] differences between 'outsourcing', 'offshore outsourcing', 'off-shoring' and 'subcontracting'. In order to "complicate" it, other authors [21] also associate the term "Body shopping" with the concept of outsourcing, i.e. staffing and professional services, usually by placing workers onsite at the client's base. The term staffing, which is also known as 'body shopping', means a contracted employee who works on the client's site or under the client's management.

The most recent decade has seen an incredible increase in outsourcing of services and has changed the way business is done. Costs of human resources, product services and "benefits perceptions" have been the primary drivers for the development in this segment. Discovering its balance first in programming it has spread to other operational ranges like IT-empowered administrations with a noteworthy part in helping administrations. This further provided a catalyst for business to process outsourcing (BPO) as a suitable suggestion in non-IT territories-e.g., routine and non-center operations in bookkeeping and money management, HR organizations, and so on. Organizations are directed to core business and they release themselves from secondary areas of business, and because of that dependence on external/outside providers will increase [1]. Dolgui and Proth [2] stated some facts about this concept. Since the seventies, outsourcing activities have been concerned about low value added products such as textiles, consumer electronics and assembly of cars. Later more outsourcing activities have been concerned about products with high added value such as software and medical equipment. Outsourcing benefits are well documented and supported by many scholars. The literature is replete with researches revealing that many organizations make this strategic decision for many reasons such as cost savings, competitive advantage, and even the opportunity to explore and take advantage of available resources. According to some authors [17], [1] the consulting sector is a trigger for the management outsourcing. However, cost effectiveness is only one of many factors that lead to the prevalence of outsourcing as an HR strategy [22]. Kang et al. [23] confirmed that many firms expressed dissatisfaction with outsourcing outcomes. In a case study of five multi-national corporations consisting of senior executives "including CEOs, purchasing managers, production managers, and outsourcing specialists" the authors concluded that in order to achieve the desired KPI's for outsourcing, different outsourcing strategies are necessary to allow different process controls [24] and that is not easy.

2 Methodology

The current section describes the approaches followed to answer the central question of the work:

H1: Will consulting firms be more profitable than outsourcing companies in a near future?

To validate this hypothesis, it has been extracted statistical information from the European database Amadeus (A database of comparable financial information for public and private companies across Europe) database [25] which contains comprehensive data on around 21 million companies across Europe. The data were collected in June 2017 for the years between 2014 and 2017, since it's the only period from which we can guarantee the existence of statistical information relevant to our analysis. The information filtered for Consulting companies has been: 1- NACE Rev. 2 (Primary codes only): 7022 - Business and other management consultancy activities; 2 - Ultimate Owners: Global; Def. of the UO: min. path of 50.01%, known or unknown shareholder; 3 - Company name:

names matching 'consulting' including branches and previous, trademarks and also known as names and exclusion of companies with no recent financial data.
Regarding Outsourcing the information filtered has been 1- NACE Rev. 2 (Primary codes only): 7830 – Other human resources provision; 2 - Ultimate Owners: Global; Def. of the UO: min. path of 50.01%, known or unknown shareholder; 3 - Company name: names matching 'outsourc' including branches and previous, trademarks and also known as names and exclusion of companies with no recent financial data. After applying the filters, all the companies were joined into two files, outsourcing and consulting companies. From this information, it has been selected the companies with EBITDA margin information (%) between 2014 and 2017. Since most of the selected companies didn´t have statistical information for all years with was only selected the companies with all the available information and with the highest rate of EBITDA margin between the three years. In the case of Consulting the selected ones were: Bure Equity AB; Ashmore Group PLC and Liberty Global PLC. From the available Outsourcing it has selected Benefit Work Szolgáltató És Tanácsadó Korlátolt Felelősségű Társaság, G4S Government and Outsourcing Services (UK) Limited and London Wall Outsourcing Investments Limited. From the information available, it has been chosen the highest three Outsourcing and Consulting companies with EBIDTA margin information between 2014 and 2017. The EBITDA analysis is important since it provides a snapshot of short-term operational efficiency and it's a good comparison mean to companies with different capital investment.

3 Data Analysis

The data extracted from the Amadeus has been filtered, previously, in order to match the needs for data analysis. To allow a comparison analysis between the outsourcing and consulting companies it has been analyzed the EBITDA margin as mentioned before. The time series evolution based on forecasting models has been used to compare both strategic companies in a near future. The forecast algorithms are used when its needed to estimate the middle range of possible values that random variables could take with relatively high probability [26]. The forecast contains embed prediction intervals for future values with 95% of probability.
From the bunch of forecasting algorithms, the Exponential Smoothing or ETS has been used since the idea behind this algorithm is to assign exponentially declining weight to observation as they go back in age, which means, recent observations have a larger weight than the old ones [27]. This algorithm, based on Holt-Winters Method [28], calculates or predicts a future value based on existing (historical) values by using the AAA (additive models of residuals, trend, and seasonality) version of the Exponential Smoothing (ETS) algorithm.

Following the initial analysis, this study analyzes each one of the Consulting or Outsourcing companies it will be more profitable in a near future, namely until 2023 based on forecast predictions.

From the outsourcing companies the forecasting values are presented on table 1, 2 and 3. From the analyzed data the three companies present an overall and growing EBITDA margin until 2023. Regarding the lower and upper confidence bound with a margin of 95% of confidence, based on optimistic and pessimist scenarios, the variation its smooth which gives us a relative confidence. The forecasting values are dependent on the previous results. The highest forecasting results from the three companies is G4S Government with the 160% of EBITDA margin in 2023. The lower and upper confidence bound for this company on this year has a variation of 29,80%. The remaining companies have also reached a positive increment. More moderate on Benefit Work (table1) and more representative, with higher variation, the company London Wall (table 3).

Table 1 – Benefit Work

Date Time	EBITDA Margin (%)	EBITDA Margin Forecast (%)	Lower Confidence Bound (%)	Upper Confidence Bound (%)
2014	32,90	-	-	-
2015	57,44	-	-	-
2016	58,27	-	-	-
2017	52,30	-	-	-
2018	-	59,33	52,87	65,78
2019	-	59,96	53,50	66,41
2020	-	60,59	54,13	67,04
2021	-	61,21	54,76	67,67
2022	-	61,84	55,39	68,30
2023	-	62,47	56,02	68,93

Table 2 – G4S Government

Date Time	EBITDA Margin (%)	EBITDA Margin Forecast (%)	Lower Confidence Bound (%)	Upper Confidence Bound (%)
2014	15,3	-	-	-
2015	7,34	-	-	-
2016	53,8	-	-	-
2017	61	-	-	-
2018	-	79,62	61	61
2019	-	95,75	52,82	106,42
2020	-	111,87	68,42	123,07

2021	-	128,00	83,40	140,34
2022	-	144,12	97,61	158,39
2023	-	160,25	110,95	177,29

Table 3 – London Wall

Date Time	EBITDA Margin (%)	EBITDA Margin Forecast (%)	Lower Confidence Bound (%)	Upper Confidence Bound (%)
2014	25,8	-	-	-
2015	31,67	-	-	-
2016	49,45	-	-	-
2017	54,2	-	-	-
2018	-	66,54	60,82	72,26
2019	-	76,73	71,00	82,45
2020	-	86,91	81,19	92,64
2021	-	97,10	91,37	102,82
2022	-	107,28	101,56	113,01
2023	-	117,47	111,75	123,19

For consulting companies the forecasting values presented on table 4, 5 and 6 have showed high ratios of growing on Bure Equity (table 4) and Ashmore (table 5). Although these two companies have a gross margin the Liberty Global (Table 6) have showed the opposite with a reduced margin each year until a 22,63 EBITDA margin with a lower confidence bound of 0,18 and an upper confidence bound of 48,53%. Despite the positive contend from the other two key players these values are the opposite, which is related with the negative variation of the previous years.

Table 4 – Bure Equity

Date Time	EBITDA Margin (%)	EBITDA Margin Forecast (%)	Lower Confidence Bound (%)	Upper Confidence Bound (%)
2014	20,5	-	-	-
2015	19,25	-	-	-
2016	16,25	-	-	-
2017	72,46	-	-	-
2018	-	77,64	38,16	117,12
2019	-	94,13	53,43	134,84
2020	-	110,63	68,73	152,53
2021	-	127,13	84,05	170,20
2022	-	143,62	99,39	187,85
2023	-	160,12	114,76	205,48

Table 5 – Ashmore Group

Date Time	EBITDA Margin (%)	EBITDA Margin Forecast (%)	Lower Confidence Bound (%)	Upper Confidence Bound (%)
2014	12	-	-	-
2015	14,62	-	-	-
2016	21,12	-	-	-
2017	60,54	-	-	-
2018	-	69,83	45,11	94,55
2019	-	85,78	60,30	111,27
2020	-	101,73	75,49	127,97
2021	-	117,68	90,71	144,65
2022	-	133,63	105,94	161,32
2023	-	149,58	121,18	177,98

Table 6 – Liberty Global

Date Time	EBITDA Margin (%)	EBITDA Margin Forecast (%)	Lower Confidence Bound (%)	Upper Confidence Bound (%)
2014	53,2	-	-	-
2015	64,91	-	-	-
2016	60,31	-	-	-
2017	42,99	-	-	-
2018	-	42,66	23,12	62,20
2019	-	38,66	18,51	58,80
2020	-	34,65	13,91	55,39
2021	-	30,64	9,32	51,96
2022	-	26,64	4,75	48,53
2023	-	22,63	0,18	45,08

4 Discussion

In the analysis of the results all the available indicators were considered by the database used and a conclusion was made based on the most used indicators in FEACO reports [4]. According to tables 1,2 and 3 for the outsourcing companies the EBITDA margin are increasing every year with an annual variation average of 0.03% for Benefit Work, 0.17% for G4S and 0.13% for London Wall respectively. By the opposite, the annual average variation from consulting companies between 2017 and 2023 was 0.14% for Bure, 0.16% for Ashmore and a negative variation of -0,1 for Liberty since the previous observed values have being decreasing along the years where data was available. From these analysis, there are no such difference between the outsourcing and consulting companies, specially with the data used which can´t gives us an overview one of the most well-known indicators used in the valuation of companies, the EBITDA, which gives the possibility of not only analyzing the final result of the organization, but also the process as a whole is able to measure the productivity and efficiency of the company. The final forecast result for the last year, 2023, indicated a higher % of EBITDA margin for Consulting companies excepted the Liberty (Table 6). From the Outsourcing companies the margin of forecast have been increasing slightly (table 1) and higher on table 2 and 3 but without the average obtained by the consulting companies.

5 Conclusions and directions for future studies

This article aimed to compare financial data between service companies of Management Consulting and Management Outsourcing. The distinction between the two is not clear in the literature nor in the tools used which shows yet the weakness of studying this subject deeply [18]. The lack of several completed indicators for a different time series, such as monthly data was an obstacle to obtain better forecasting results or even validate it through mean square error or root mean square error. The work begins at European level and several countries have attempted that via "official statistics" from central statistical bureaus or via associations. Then FEACO [8] has undertaken the task of collating the national data and even analyzing the trends across nations with a visible degree of success over the times. In our study and after typing in a set of companies with financial data in Europe several conclusions can be drawn for the 2 types of companies studied: consulting and outsourcing. The indicators used by FEACO [14] and by some authors [29] clearly stand out the consulting companies as having more attractive financial ratios. The "EBITDA" allows us to draw the conclusion that projects are not sold for hours but rather for value [3], which makes the value chain of this type of company more profitable. We also evaluated companies that provide both services and in the indicators mentioned above also these companies present better financial ratios than outsourcing firms. We can also associate worse financial performances of outsourcing companies by the prices practiced. This point is referred by Hutzschenreuter, Lewin, and Ressler [30] as well as other researchers who have surveyed insourcing and outsourcing phenomenon's [31-32, 33]. They often concluded that the cost savings or cost reductions are one of the major reasons for outsourcing or offshoring products and services. Only based on financial information, we can conclude that the work model of outsourcing firms should be reviewed or look at a mixed model under penalty of not being as profitable as it could, but we cannot consider it incorrect since it presents positive results.

6 References

[1] Nélson Santos Antonio (2012) The 'outsourcing'and management consulting in a dimensional systemic model *African J. Bus. Manag.*, vol. 6, no. 31

[2] A. Dolgui and J. Proth (2010), Supply chain engineering: useful methods and techniques

[3] M. Kubr (2002) Management consulting: A guide to the profession

[4] FEACO (2010) Survey of the European Management Consultancy 2010/2011, *Eur. Fed. Manag. Consult. Assoc*

[5] B. Ahmad (2012) Management: Evolution and Application in Our Environment., *J. Manag. Sci*

[6] R. David (2001) The Emergence and Evolution of an 'Expert' Field, *Unpubl. Ph. D. Diss. Cornell Univ*

[7] S. Canback (1998), "The Logic of Management Consulting, (Part One)," *J. Manag. Consult.*, vol. 10, no. 2, pp. 3–11

[8] FEACO (2015), "Survey of the European Management Consultancy," *Eur. Fed. Manag. Consult. Assoc.*, no. December

[9] R. L. Costa, N. Antonio, I. Miguel, and F. Martinho (2014) The Analysis of the Four Paradigms of Business Management consulting in Portugal in the Light of the Research Field of Strategy-as-practice *Adv. Manag. Appl. Econ.*, vol. 4, no. 3, pp. 39–54

[10] M. Kipping and I. Kirkpatrick (2013) Alternative pathways of change in professional services firms: The case of management consulting," *J. Manag. Stud.*, vol. 50, no. 5, pp. 777–807

[11] A. Corsi (2010) Why Management Consultants ? Between Functionalist and Critical Perspectives *Master Thesis*, no. January

[12] N. Butler (2009) What is management consultancy?

[13] A. Gross and J. Poor (2008) The global management consulting sector *Bus. Econ*

[14] FEACO (2015) Survey of the European Management Consultancy *Eur. Fed. Manag. Consult. Assoc*

[15] R. Fincham, T. Clark, and K. Handley (2008) Configuring expert knowledge: the consultant as sector specialist *J*

[16] C. Teixeira and L. Pereira (2015) Pereira Diamond: benefits management framework *Int. J. Bus*

[17] J. Dibbern, T. Goles, R. Hirschheim, and B. Jayatilaka (2004) Information systems outsourcing: a survey and analysis of the literature *ACM Sigmis Database*

[18] M. Belcourt (2006) Outsourcing - The benefits and the risks *Hum. Resour. Manag. Rev.*, vol. 16, no. 2, pp. 269–279

[19] D. Simchi-Levi, P. Kaminsky, and E. Simchi-Levi (2004) Managing The Supply Chain: Definitive Guide

[20] A. Dolgui and J.-M. Proth (2013) Outsourcing: definitions and analysis *Int. J. Prod. Res.*, vol. 51, no. January 2015, pp. 6769–6777

[21] V. Weerakkody and Z. Irani (2010) "A value and risk analysis of offshore outsourcing business models: An exploratory study," *Int. J. Prod. Res.*, vol. 48, no. 2, pp. 613–634

[22] J. Sriwongwanna (2009) Understanding the Impact of Outsourcing Human Resource Activities on Employee Attitudes and Behaviours

[23] M. Kang, X. Wu, P. Hong, and Y. Park (2012) Aligning organizational control practices with competitive outsourcing performance, *J. Bus. Res*

[24] A. H. Yagoob and Z. Ting (2015), "Application of Strategic Sourcing Practice in Public and Private Sectors : Literature Review," vol. 7, no. 24, pp. 40–52

[25] "Amadeus - SABI - Bureau van Dijk database. Retrieved from https://www.bvdinfo.com/en-gb/home," 2018. [Online]. Available: https://www.bvdinfo.com/en-gb/home.

[26] R. Hyndman and G. Athanasopoulos (2018) Forescasting: Principles and Practice, Monash University, Australia. 2nd Edition

[27] B. Iqelan (2009), Comparison of Parametric and Nonparametric Techniques for Water Consumption Forecasting, International Journal of Scientific & Engineering Research, Volume 8, Issue 1

[28] A. Shahin (2016) Using Multiple Seasonal Holt-Winters Exponential Smoothing to Predict Cloud Resource Provisioning, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 11

[29] N. Capon, J. U. Farley, and S. Hoenig (2011) Determinants of Financial Performance *Manag. Sci. vol 36 no 10*, vol. 36, no. 10, pp. 1143–1159

[30] T. Hutzschenreuter, A. Y. Lewin, and W. Ressler (2011) The growth of white-collar offshoring: Germany and the US from 1980 to 2006, *Eur. Manag. J.*

[31] J. Freeman and M. Minow (2009), Government by contract: Outsourcing and American democracy

[32] S. Mukherji and J. Ramachandran (2007) Outsourcing: practice in search of a theory *IIMB Manag. Rev*

[33] C. Williams (2009) Rationales for outsourcing domestic services to off-the-books workers, *J. Econ. Stud.*

FPGA-based accelerator design for Echo-State networks

Josep L. Rosselló *, M. L. Alomar, Erik S. Skibinsky-Gitlin, Christiam F. Frasser, Vincent Canals, Eugeni Isern, Fabio Galán-Prado, Alejandro Morán, and Miquel Roca

Physics Department, Universitat de les Illes Balears
Cra. Valldeomssa km. 7.5, Palma de Mallorca, Balears, Spain
j.rossello@uib.es

Abstract. The hardware implementation of Echo State Networks (ESN) can be applied to situations where a quick response is needed in relation to how a certain signal will evolve. This is due to the possibility of connecting the ESN's neurons in parallel, which accelerates the calculation process considerably. In this article, we present a proposal for the compact implementation of ESN in Field Programmable Gate Arrays (FPGAs). To maximize the number of neurons, the synapses are implemented using adders and multiplexers instead of multipliers. The hardware implementation has been tested for the prediction of the Santa Fe time series dataset. The proposed approach allows for significant savings in terms of energy, hardware resources and computing time compared to other recently published solutions.

Keywords: Echo State Networks, FPGA, Time series forecasting

1 Introduction

Numerous applications require the use of specific hardware to implement machine-learning applications. Those systems take advantage of the inherent parallelism in the neural processing, that may be beneficial in terms of speed, power, reliability and cost [1, 2]. As an example, hardware neural networks (HNNs) are necessary for high-volume processing and real-time applications, such as image search and data mining [3, 4].

A great research effort has been made to develop efficient HNN implementations [5–8]. However, the synapses' implementation constrains the viability of implementing massive networks in a single chip. For these reasons, the use of

* This work has been partially supported by the Spanish Ministry of Economy and Competitiveness (MINECO), the Regional European Development Funds (FEDER), and the Comunitat Autnoma de les Illes Balears under grant contracts TEC2014-56244-R, TEC2017-84877-R and a fellowship (FPI/1513/2012) financed by the European Social Fund (ESF) and the Govern de les Illes Balears (Conselleria d'Educació, Cultura i Universitats)

approximate multipliers [9, 10] has been proposed to reduce hardware at the cost of accuracy loss.

Echo State Networks (ESN) is suited to time-series prediction and classification tasks [11] and has been successfully applied in numerous domains, such as robot control [12], image/video processing [13], or financial forecasting [14]. Therefore, fast hardware designs implementing ESN systems [15, 16] is of interest for these applications, which require real-time intensive data processing [17].

In this article, we present a proposal for ESN hardware implementation using few hardware resources. The proposed design presents low power characteristics. As a demonstration of the validity of the approach, we implement a large reservoir network within an FPGA and evaluate its performance for a traditional benchmark on time-series processing (Santa-Fe prediction task). The results are compared with some previously-published works.

2 Methodology

2.1 Echo State Networks

ESN differs from Recurrent Neural Networks in that the synaptic weights between neurons are kept fixed and only the connections from the network to a measurement output layer are modified by learning. This reduces the training to a classical linear regression problem. The architecture of a reservoir computing system consists of a total of N internal processing nodes (the neurons) each one providing a given value $x_{k,i}$, where $i \in \{1, 2, .., N\}$ is the neuron index, k represents the evolution during time ($k \in \{1, 2, ..., L\}$) and L is the total number of samples taken from the reservoir. Therefore, the time evolution of internal nodes of the reservoir is described by a matrix with L rows and N columns $\mathbf{X}$ (the design matrix). The state of the network at a given time k is defined by the kth row of the design matrix $\mathbf{x}(k)$ and the time evolution of a given node is stored in the ith column $\mathbf{x}_i$. The output response of the reservoir is computed in two phases. First, the current reservoir state [$\mathbf{x}(k)$] is updated according to a nonlinear function of the weighted sum of the neuron inputs [M external time-dependent inputs $\mathbf{u}(k) = (u_1(k), u_2(k), ..., u_M(k))$], and N internal ones coming from the reservoir's neurons evaluated in the previous time step, $\mathbf{x}(k-1)$ following the expression:

$$\mathbf{x}(k) = f\left[\mathbf{W}_{in}\mathbf{u}(k) + \mathbf{W}\mathbf{x}(k-1)\right] \tag{1}$$

where f is the activation function $f : \mathbb{R}^N \to \mathbb{R}^N$, $\mathbf{W}_{in}$ and $\mathbf{W}$ are two $N \times M$ and $N \times N$ weight matrices respectively. The network is used to evaluate a total of Q outputs [$\hat{\mathbf{y}}(k) = (\hat{y}_1(k), \hat{y}_2(k), ..., \hat{y}_Q(k))$] that are obtained preforming linear combinations of the reservoir states:

$$\hat{\mathbf{y}}(k) = \mathbf{x}^T(k)\mathbf{W}_{out} \tag{2}$$

where $\mathbf{W}_{out}$ is a $N \times Q$ weight matrix obtained using a linear regression with respect the expected outputs [$\mathbf{y}(k) = (y_1(k), y_2(k), ..., y_Q(k))$].

2.2 Training method

Assuming we can take a total of L measurements for $\mathbf{x}(k)$ and $\mathbf{y}(k)$, we define $\mathbf{Y}$ as the feature matrix of $L \times Q$ that will be approximated by the network (composed of L row vectors of Q elements $\mathbf{y}(k)$). Then we have that $\mathbf{W}_{out}$ is estimated using the Moore-Penrose pseudo-inverse:

$$\mathbf{W}_{out} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} \tag{3}$$

In the reservoir computing scheme, matrices $\mathbf{W}_{in}$ and $\mathbf{W}$ are taken fixed while $\mathbf{W}_{out}$ is conveniently trained using expression 3 or other similar linear fitting.

2.3 Low cost hardware implementation Echo State Networks

The number of neurons employed in ESNs is usually high (typically between 50 and 1000, although some applications require much larger networks to achieve the desired accuracy [13]), which makes particularly challenging the hardware implementation of these systems. Given the large number of products to be implemented due to the high number of synapses, multipliers expend a significant portion of the integrated circuit resources. We limit the possible weights to integer powers of two and sums of powers of two so that shift registers can be employed instead of multipliers.

For an standard ANN implementation that use back-propagation as learning algorithm, the constraint on the weights is not desirable since it leads to lower network performance [18]. Nonetheless, we show that for ESN with fixed connections, the proposed approach only implies a minor accuracy loss. For the estimation of the output $[\hat{\mathbf{y}}(k)]$ from expression (2) we use the dedicated embedded multipliers integrated in the FPGA device (DSP blocks). The overall area impact of this computation inside the FPGA is relatively low since the logical elements needed to implement the neural network are not used and the number of multipliers at the output layer is significantly lower than the synapses' multipliers (that would limit the maximum neural fan-in). Regarding the network topology, a simple cyclic architecture presents a similar performance to the classical random one while it minimizes the number of connections [19] and optimize the packing efficiency. In Fig.1 we show the case of a simple cyclic reservoir (SCR) with one input ($u(t)$) and one output ($Q = 1$). For simplicity, the connections between internal units have the same weight r whereas the inputs are connected to the reservoir with a weight that is positive $v =\mid v \mid$ or negative $v = - \mid v \mid$ with the same probability and same absolute value. To improve the experimental results, parameters r and v are firstly adjusted numerically to find the optimum weight configuration.

Fig.2a illustrates a general circuit design for a two-input sigmoid neuron necessary to build the cyclic reservoir when only one input signal is processed. The fixed-point two's complement notation is assumed for all signals so that both positive and negative values can be represented. The first neuron's input

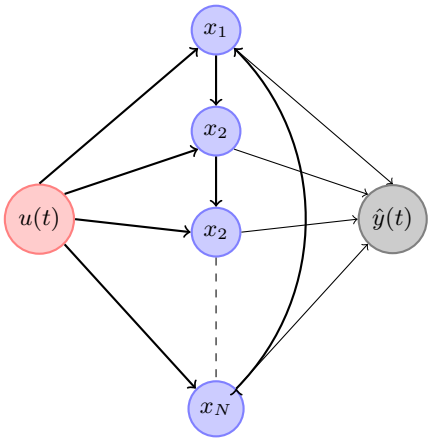


Fig. 1. The cyclic Echo State Network scheme is composed of three blocks: Input layer, cyclic neural network and the output.

$[u(t)]$ refers to the external input signal (to be processed by the network) and the second one $[x_{i-1}(t-1)]$ to the state of a neighboring neuron evaluated at the previous time step. A resolution of n bits is considered for the input and of m bits for the weights (v and r). The multiplier's output is truncated to n bits taking the most significant of the result, but a higher or lower resolution could be employed depending on the desired accuracy.

A simple piece-wise linear approximation with three segments [5] is used for the implementation of the activation function, due to its simple implementation. More accurate designs (e.g., [6, 7]) could be employed to improve the network's performance at the cost of higher hardware requirements. The scheme of Fig.2a can be simplified to that of Fig.2b when the weight resolution is limited to a few bits, more specifically $m = 4$. In this case, the full multipliers can be substituted by shift-and-add blocks. Such "multiplier-less" approach enable great hardware saving at the cost of constraining the possible values of the connection weights. The shift-and-add block is depicted in Fig.2c. Basically, it performs a multiplication of the input signal $[u(t)]$ by the corresponding weight (v) with a pair of shift registers and an adder. Some additional circuitry is included to perform the negation of the shifted values in case it is necessary. A multiplexer is employed to provide either the number that directly results from the shift register or its corresponding negative value depending on a selection signal. A decoder configures the shift registers (with the number of required shifts, $sh1$ and $sh2$) and controls the activation of the negations ($neg1$ and $neg2$) as a function of the weight value (v). By way of example, a single right shift of the input ($sh1 = 1$) performs a multiplication by 0.5 while two shifts ($sh2 = 2$) are equal to a factor of 0.25. The direct addition (with $neg1 = neg2 = 0$, indicating that no negation of the shifted values is necessary) of these two shifted magnitudes results in a weight $v = 0.75$. The weight value $v = 0.875$ can be implemented by

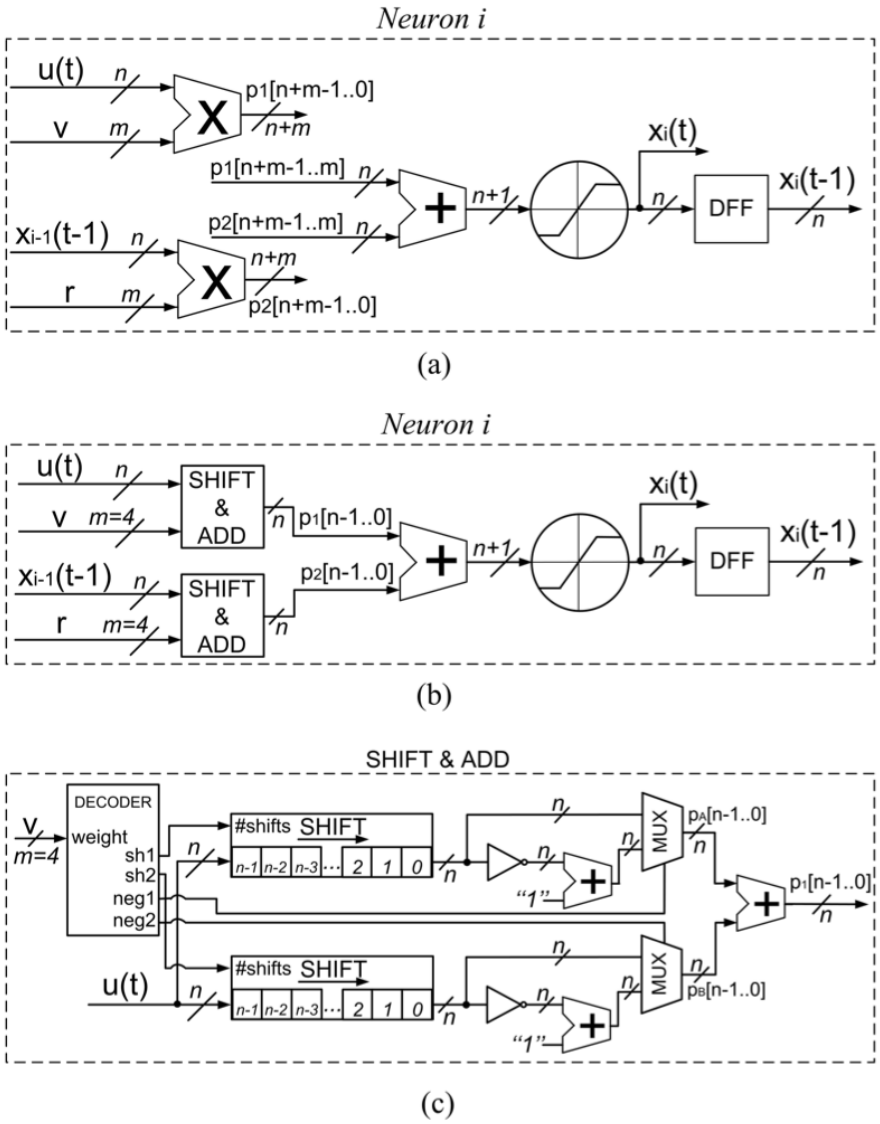


Fig. 2. (a-c) Neuron design: (a) general circuit design of the neuron; (b) reduced implementation scheme when the weight resolution is limited to few bits ($m = 4$) and the multipliers are replaced by simple shift-and-add blocks; (c) description of the shift-and-add block

selecting no shifts and no negation for the first shift register ($sh1 = 0$, $neg1 = 0$) and three shifts with a negated output for the second one ($sh2 = 3$, $neg2 = 1$) so that the input signal $u(t)$ is weighted by the factor $v = 1 - 0.125 = 0.875$. A negative factor, for instance $v = -0.5$, may be obtained through the negation of both shifted magnitudes ($neg1 = neg2 = 1$), where each one is obtained with two displacements ($sh1 = sh2 = 2$) so that $v = -0.25 + (-0.25) = -0.5$. In the case it is desired a generic design, the circuit of Fig.2c can be used. For FPGA implementations in which the training of r and v are done off-line, an ad-hoc design implying the implementation of the exact shifts for each neuron input can be used, further simplifying the hardware and increasing the processing speed. The transfer function f in eq. (1) determines the behavior of the analog (discrete-time) neuron. For the special case of this work, in which the training is performed at the output layer and not implementing a back-propagation algorithm, a piece-wise linear function is able to provide very good fitting results along with a compact hardware implementation. In Fig.3 we show the non-linear function used for each internal neuron of the reservoir. The function is easily reproduced in hardware using a few gates and a multiplexer.

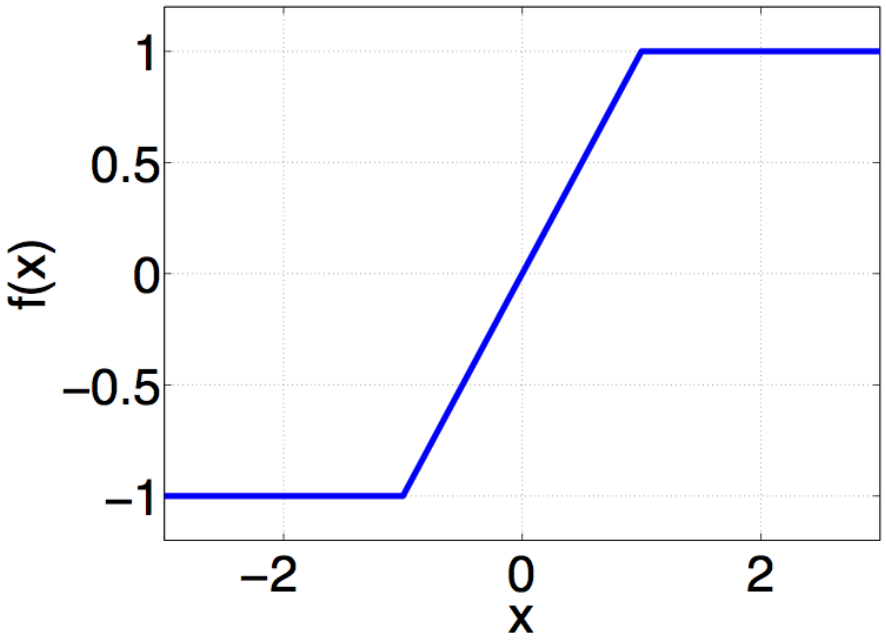


Fig. 3. Piece-wise function used as non-linear activation function

2.4 Benchmark prediction Task

In time-series prediction or forecasting the objective is to predict future values based on previously observed ones. Thus, the input sequences are mapped onto a real-valued output sequence that represents one-step or several-step ahead predictions of the desired variable. That is, the value of the series at the current time is introduced each time step as input to the system and the time-series value corresponding to the next (or several) time step must be predicted. In this work we test the proposed ESN hardware with respect to a widely used benchmark that is the Santa Fe [20] time series prediction task. The time series processing is divided in two steps:

- Part of the time-series is used for off-line training using an R script. The optimum v and r values are selected along with the output weights ($\mathbf{W}_{out}$). Then, the network is automatically generated using a hardware description language code (VHDL).
- The rest of the time-series is digitized to 16bits two's complement that is transferred to the on-chip RAM memory of the FPGA for its processing.

The typical processing task is to predict the next sample of the time series before it has been injected into the reservoir computer (one-step ahead prediction). The performance of this task is evaluated using the normalized mean square error (NMSE):

$$NMSE = \frac{\sum_{i=1}^{L} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{L} \left(y_i - \bar{y}\right)^2} \tag{4}$$

where $\mathbf{y} = \big(y(1), y(2), .., y(L)\big)$ is the time series to be predicted (target), $\hat{\mathbf{y}} = \big(\hat{y}(1), \hat{y}(2), .., \hat{y}(L)\big)$ is the predicted value provided by the output layer of the reservoir following equation (2), and $\bar{y}$ is the mean value of $\mathbf{y}$. Parameter L is the number of samples.

The VHDL code is automatically generated from an R script and is composed of three parts: RAM memory (containing the input to be processed), the cyclic reservoir (constructed using the programmable logic elements of the FPGA), and the output layer (using the dedicated multipliers of the FPGA). The code is used to configure an ALTERA Cyclone IV (EP4CE22F17C6N) FPGA chip.

The Santa Fe prediction task A representative example is the Santa Fe laser time-series prediction task, a widely used benchmark [19]. The task consists in forecasting an experimental recording of the output power of a far-infrared laser operating in the chaotic regime. It is usually evaluated for one-step ahead predictions. Data is available at [20]. In this work, we employ 4000 samples of the original laser dataset, the first 75% for training and the remaining 25% for testing.

The goal for the Santa-Fe task is to predict the next sample in the chaotic time trace before it has been injected into the reservoir computer (one-step ahead prediction).

3 Results

We compare the performance of the proposed model with respect to different widely used benchmark tasks and comparing with some previously published works. Regarding the area impact of the proposed methodology, it represents an 86% of area reduction if compared with the standard digital realization when two inputs for each neuron are used and therefore the number of synapses' multipliers are reduced to the minimum in the conventional digital approach.

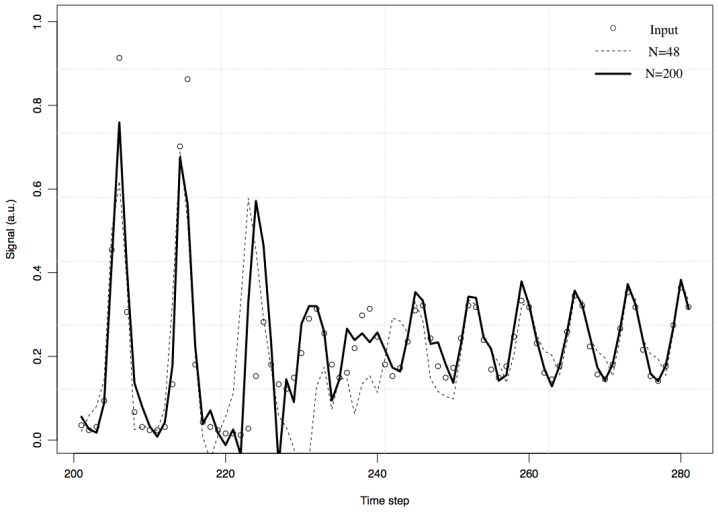


Fig. 4. Fragment of the Santa Fe time-series test set: original values and one-step ahead predictions performed by the proposed reservoir implementation with $N = 48$ and $N = 200$ neurons. As can be appreciated, the network is able to adapt to the abrupt changes of the input

The proposed methodology is used to synthesize echo state networks with cyclic topology (SCR) (Fig.1), encoded using VHDL on an ALTERA Cyclone IV FPGA. The performance of the system is tested for the Santa Fe time-series prediction task [20]. Networks with different sizes with up to 200 neurons are implemented and analyzed using the proposed "multiplier-less" approach (Fig.2b) with a precision of 16 bits ($n = 16$). The training is performed following the methodology mentioned in section 2.1. A convenient numerical model of the hardware reservoir is employed for the learning phase. Finally, once the prediction error has been scanned for all the possible network configurations (values of r and v), the hardware realization is set up with the optimum weights and evaluated using the test set. The parameters providing better fitting ($r = 0.875$ and $v = 1$) are obtained and the FPGA is configured with those values. The test

set is then stored into the internal RAM memory, thus providing a new input value to the reservoir every time step (each N clock cycles). The resulting outputs (individual neuron states) are processed by the FPGA providing the output $\hat{\mathbf{y}}$ each N clock cycles. This computation is performed following equation (2) in a total of N clock cycles so that the processing speed of the proposed design is f/N, where f is the clock frequency that for all the experiments performed in this work is fixed to $50MHz$. This output is extracted from the FPGA with a logic analyzer and used to calculate the system's performance as the error between the estimated and targeted values. For this task we employ a total of 4000 samples of the original laser dataset, the first 3000 for training and the remaining 1000 for testing.

Fig.4 illustrates the experimental predictions obtained through the proposed design when using 48 and 200 neurons. As can be observed, the fitting of the measurements taken in the FPGA is improved when increassing N.

Table 1. Performance results of the Santa Fe experiments for the proposed design and some previously published models

Processing	Ref.	N	NMSE (Hard.)	NMSE (Soft.)	speed (pps)	Power (W)	PDP (μJ)
Optoelectronic	[21]	200	-	0.02	-	-	-
Optoelectronic	[22, 23]	388	0.106	-	$1.3 \cdot 10^7$	150	11.5
Optoelectronic	[23]	400	-	0.022	-	-	-
Software	[19]	200	-	0.009	-	-	-
Software	[19]	50	-	0.0196	-	-	-
Software	This work	200	-	0.0766	-	-	
FPGA	This work	200	0.079	-	$2.5 \cdot 10^5$	< 1.5	6

In Table 1 we show the performance of the proposed model measured in terms of the NMSE, speed (in points predicted per second) and power dissipation for the processing of the Santa-Fe time series prediction task. Comparison with previously published models is also included in the table. We distinguish between theoretical values of NMSE obtained from high-precision numerical calculations (Soft.) and NMSE values obtained from experimental settings (Hard.). These two values can differ significantly due to the intrinsic complexity of experimental settings that may present both system and quantization noise. This is evidenced in references [22, 23], where the expected NMSE provided by software is considerably lower than the measured one. We also show the results of different studies that are purely numerical as the work in [19] showing the expected performance of Simple Cycle Reservoir designs or the paper in [21] where a semiconductor ring laser with optical feedback (SRL) is numerically simulated. We also provide in the table the power-delay product (PDP) achieved by the experimental settings (a classical figure of merit of the overall hardware performance).

As can be appreciated the proposed design is able to provide a factor of two of lower PDP when compared with [22, 23].

4 Conclusions

In this work, we have presented a digital implementation of ESN. It is shown that the connection weights can be limited to a few discrete values without compromising the system's performance. The validity of the resulting implementation has been demonstrated for the Santa Fe time-series prediction task. Performance comparisons with previous works are shown showing that the proposed model present competitive results. Those comparisons show that the proposed model represent a considerable improvement in terms of speed and power dissipation while is able to provide a similar accuracy than previous models. The proposed solution allows the parallel computation of all the recurrent network with a processing speed of the order of MHz and therefore it is suited to supporting real-time signal processing applications. On the other hand, the proposed approach can be useful to perform specialized systems implementing computational intelligence techniques that requires low power consumption. Potential applications include speech recognition [24], robotics [12], wireless sensor networks [25], predictive controllers [26] and the classification of medical signals [27] among others. To summarize, it has been shown that the use of low-resolution weights to implement the internal synapses of the neurons has little effect on the system's performance while it allows a considerable reduction of the hardware since the use of binary multipliers are avoided. This observation makes possible a very compact implementation of massive reservoir networks with parallel processing capabilities.

References

1. Misra, J., Saha, I.: Artificial neural networks in hardware: A survey of two decades of progress. Neurocomputing **74**(1-3) (2010) 239–255
2. Baptista, F.D., Morgado-Dias, F.: Automatic general-purpose neural hardware generator. Neural Computing and Applications **28**(1) (2017) 25–36
3. Krizhevsky, A., Sutskever, I., Geoffrey E., H.: ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS2012) (2012) 1–9
4. Morro, A., Canals, V., Oliver, A., Alomar, M.L., Galan-Prado, F., Ballester, P.J., Rossello, J.L.: A Stochastic Spiking Neural Network for Virtual Screening (2017)
5. Basterretxea, K., Tarela, J.M., del Campo, I.: Digital design of sigmoid approximator for artificial neural networks. Electronics Letters **38**(1) (2002) 35–37
6. Baptista, D., Morgado-Dias, F.: Low-resource hardware implementation of the hyperbolic tangent for artificial neural networks. Neural Computing and Applications **23**(3-4) (2013) 601–607
7. Nascimento, I., Jardim, R., Morgado-Dias, F.: A new solution to the hyperbolic tangent implementation in hardware: Polynomial modeling of the fractional exponential part. Neural Computing and Applications **23**(2) (2013) 363–369

8. Carrasco-Robles, M., Serrano, L.: Accurate differential tanh(nx) implementation. International Journal of Circuit Theory and Applications **37**(5) (2009) 613–629
9. Lotrič, U., Bulić, P.: Applicability of approximate multipliers in hardware neural networks. Neurocomputing **96** (2012) 57–65
10. Nedjah, N., De MacEdo Mourelle, L.: Reconfigurable hardware for neural networks: Binary versus stochastic. Neural Computing and Applications **16**(3) (2007) 249–255
11. Lukoševičius, M., Jaeger, H., Schrauwen, B.: Reservoir Computing Trends. KI - Künstliche Intelligenz **26**(4) (2012) 365–371
12. Antonelo, E.A., Schrauwen, B.: On Learning Navigation Behaviors for Small Mobile Robots with Reservoir Computing Architectures. IEEE Transactions on Neural Networks and Learning Systems **26**(4) (2015) 763–780
13. Jalalvand, A., Wallendael, G.V., Walle, R.V.D.: Real-Time Reservoir Computing Network-Based Systems for Detection Tasks on Visual Contents. In: Proceedings - 7th International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2015. (2015) 146–151
14. Lin, X., Yang, Z., Song, Y.: Short-term stock price prediction based on echo state networks. Expert Systems with Applications **36**(3 PART 2) (2009) 7313–7317
15. Alomar, M.L., Soriano, M.C., Escalona-Morán, M., Canals, V., Fischer, I., Mirasso, C.R., Rosselló, J.L.: Digital Implementation of a Single Dynamical Node Reservoir Computer. IEEE Transactions on Circuits and Systems II: Express Briefs **62**(10) (2015) 977–981
16. Alomar, M.L., Canals, V., Perez-Mora, N., Martínez-Moll, V., Rosselló, J.L.: FPGA-based stochastic echo state networks for time-series forecasting. Computational Intelligence and Neuroscience **2016** (2016)
17. Rossello, J.L., Alomar, M.L., Morro, A., Oliver, A., Canals, V.: High-Density Liquid-State Machine Circuitry for Time-Series Forecasting. International Journal of Neural Systems **26**(5) (2016) 1550036
18. Marchesi, M., Orlandi, G., Piazza, F., Uncini, A.: Fast Neural Networks Without Multipliers. IEEE Transactions on Neural Networks **4**(1) (1993) 53–62
19. Rodan, A., Tiňo, P.: Minimum complexity echo state network. IEEE Transactions on Neural Networks **22**(1) (2011) 131–144
20. Weigend, A.S., Gershenfeld, N.A.: Results of the time series prediction competition at the Santa Fe Institute. In: IEEE International Conference on Neural Networks - Conference Proceedings. Volume 1993-Janua. (1993) 1786–1793
21. Modeste Nguimdo, R., Verschaffelt, G., Danckaert, J., Van Der Sande, G.: Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback. IEEE Transactions on Neural Networks and Learning Systems **26**(12) (2015) 3301–3307
22. Brunner, D., Soriano, M.C., Mirasso, C.R., Fischer, I.: Parallel photonic information processing at gigabyte per second data rates using transient states. Nature Communications **4** (2013)
23. Hicke, K., Escalona-Morán, M., Brunner, D., Soriano, M.C., Fischer, I., Mirasso, C.R.: Information Processing Using Transient Dynamics of Semiconductor Lasers Subject to Delayed Feedback. IEEE Journal of Selected Topics in Quantum Electronics **19**(4) (2013) 1501610–1501610
24. Lee, M., Hwang, K., Park, J., Choi, S., Shin, S., Sung, W.: FPGA-based low-power speech recognition with recurrent neural networks. In: IEEE Workshop on Signal Processing Systems, SiPS: Design and Implementation. (2016) 230–235

25. Mathews, E., Poigné, A.: An Echo State Network based pedestrian counting system using wireless sensor networks. In: 2008 International Workshop on Intelligent Solutions in Embedded Systems (WISES 2008). (2008) 1–14
26. Li, H., Zhang, D., Foo, S.Y.: A stochastic digital implementation of a neural network controller for small wind turbine systems. IEEE Transactions on Power Electronics **21**(5) (2006) 1502–1507
27. Raghunathan, S., Gupta, S.K., Ward, M.P., Worth, R.M., Roy, K., Irazoqui, P.P.: The design and hardware implementation of a low-power real-time seizure detection algorithm. Journal of Neural Engineering **6**(5) (2009) 056005

Stacked LSTM Snapshot Ensembles for Time Series Forecasting

Sascha Krstanovic and Heiko Paulheim

University of Mannheim, Germany
Research Group Data and Web Science
sascha@informatik.uni-mannheim.de
heiko@informatik.uni-mannheim.de

Abstract. Ensembles of machine learning models have proven to improve the performance of prediction tasks in various domains. The additional computational costs for the performance increase are usually high since multiple models must be trained. Recently, snapshot ensembles [13] gained attention as they provide a comparably cheap way of ensemble learning for artificial neural networks (ANNs). We extend snapshot ensembles to the application of time series forecasting, which comprises two essential steps. First, we show that determining reasonable selections for sequence lengths can be used to efficiently escape local minima. Additionally, combining the forecasts of snapshot LSTMs with a stacking approach greatly boosts the performance compared to the mean of the forecasts as used in the original snapshot ensemble approach. We demonstrate the effectiveness of the algorithm on five real-world datasets and show that the forecasting performance of our approach is superior to conservative ensemble architectures as well as a single, highly optimized LSTM.

Keywords: Time Series, LSTM, ARIMA, Ensembles, Stacking, Meta-Learning

1 Introduction

Estimating the future development of continuous data generated by one or more signals has been an ongoing research field of interest for various applications. For example, automated financial forecasting is vital in today's markets. Further, sensor generated data driven by the Internet of Things requires robust methods for reliable forecasts of temporal data. Long Short-Term Memory (LSTM) [10] has proven to be an effective method for a variety of sequence learning tasks such as time series forecasting. Relying on a single LSTM, however, is prone to instability due to the dynamic behavior of time series data. Additionally, the optimization of LSTM parameters is a hard problem that requires time intensive fine tuning.

Another difficulty when dealing with time series problems lies in the slicing of the data, i.e., how many past values should be considered for training the model

and generating forecasts. It is common practice to determine the top periodicity using a fast fourier transformation and power spectra, and train one or more models based on that periodicity. This approach is prone to incompleteness because information may be encoded across patterns of varying periodicity in the series. It is also a time consuming task as identifying the optimal sequence length is usually part of a manual preprocessing step. For these reasons, it is a challenge to create machine learning frameworks that are able to produce automated forecasts for a given series. Even a greatly tuned model fails to find important relationships in time series data if the selected time lags can not represent these patterns. Therefore, a framework that can incorporate multiple sequence lengths is desirable.

We introduce a meta learning approach based on snapshot ensembles that provides superior and robust forecast estimates across different datasets. In contrast to the original idea of snapshot ensembles, we do not adapt the parameters of the LSTM but leave them unchanged. Instead, we use different slices of the training data in order to escape local minima and to detect time-dependent patterns. Our proposed approach enables the automated generation of time series forecasts for a given series $y_1, ..., y_n$, including preprocessing steps like data standardization, periodicity detection, data slicing and splitting. Hence, the amount of required manual work is greatly reduced by the proposed framework.

By sequentially training LSTMs with periodicities of decreasing strength, our algorithm is able to learn the different patterns of the respective seasonalities. This allows for higher generalization of the final model, thereby providing estimates that are robust with respect to the underlying data generation process.

The rest of this paper is structured as follows. Section 2 provides an overview of existing approaches to time series forecasting and their application within ensemble frameworks. In Section 3, we introduce the concept of snapshot ensembles and explain our approach for their extension to the task of time series forecasting. We show that our method outperforms previous approaches on five datasets in Section 4. Eventually, we conclude and give an outlook on future research directions in Section 5.

2 Related Work

Time series forecasting is a highly common data modeling problem since temporal data is generated in many different contexts. Classical forecasting approaches are based on autoregressive models such as ARIMA, ARIMAX, and Vector Autoregression (VAR) [7] [20]. Here, a forecast estimate is dependent on a linear combination of the past values and errors. Autoregressive models work well if the assumption of stationarity is true and the series is generated by a linear process [1]. On the other hand, these hard assumptions limit the effectiveness of autoregressive models if one deals with non linear series, as it is the case with the majority of practical time series problems.

LSTM, a particular variant of artificial recurrent neural networks (RNN), overcomes these shortcomings as it makes no assumptions about the prior dis-

tribution of the data. One can think of RNNs as regular feed-forward networks with loops in them. This enables RNNs to model data with interdependencies such as autoregression. It has been shown that artificial neural networks with one hidden layer can, in theory, approximate a continuous function arbitrarily well [11]. As the RNN gets deeper, vanishing or exploding gradients often lead to poor model performance [4] [18]. LSTMs solve this problem with a gating mechanism that controls the information flow in the neurons. LSTMs show superior performance in a variety of sequence learning tasks such as machine translation [8] [21].

Since autoregressive models perform well for linear series and neural networks for non linear data, there exist a number of hybrid approaches that make use of these characteristics. In those cases, the data is first split into a linear and a non linear component and each one is modeled independently. The individual results are then combined additively to determine the final estimate [2] [3] [22] [24].

The sequential nature of LSTMs has led to them being studied in the context of time series forecasting intensively. [6] [15] [17] describe applications of LSTMs for forecasting tasks. [1] [14] propose frameworks of LSTM ensembles with independently trained models. Finally, snapshot ensembles constitute a way to construct an ensemble of dependent ANNs at comparably low computational costs. A more detailed description is given in Sec. 3.1. We extend this method to recurrent neural networks and sequential problems.

Finding periodicities in time series data is a key part in the preprocessing of time series data and proposes a major challenge for the automation of machine generated forecasts. [5] propose a variation of the approximate string matching problem for automated periodicity detection. [26] develop strategies on diversity generation and build ensembles of the resulting models. In [27], a number of heterogeneous models are arbitrated by a meta learner. An approach that exploits fourier transformations is given in [19]. We will use a similar methodology in the course of this paper.

3 Time Series Forecasting and Snapshot Ensembles

Time series data is subject to a number of properties due to interdependencies across observations:

1. Autoregression. In contrast to a machine learning setup where observations are independent from one another, sequence learning tasks are characterized by dependencies between observations. This has effects on data sampling and model evaluation as drawing completely random subsamples is not possible. Hence, a suitable sample strategy is indispensable when modeling temporal data.
2. Structural patterns and changes. Due to trend and seasonality effects, the behavior of a time series is subject to repetition and change at the same time. While similar patters may repeat over time, the frequency and intensity of those are usually not constant. This is one reason why ensemble methods

are a powerful tool for time series data as each of the snapshot models incorporates information of different behavior.

3.1 Introduction to Snapshot Ensembles

Snapshot Ensembles propose a novel technique to obtain an ensemble of ANNs at the same computational costs as fully training a single ANN. The central idea is that instead of training a number of independent ANNs, only one ANN must be optimized. In the process of optimization, the ANN converges to a number of different local minima. Every time the ANN reaches a local minimum, the model snapshot is stored along with its architecture and weights. The final weights of a snapshot serve as the weight initialization of the succeeding snapshot LSTM. Finally, each snapshot provides a prediction estimate and the ensemble predictor is calculated as the mean of the snapshot estimates. It was shown that this combination yields advantageous performance compared to the single best estimate [13].

3.2 Extending Snapshot Ensembles to Sequence Problems

Time series forecasting can be interpreted as a sequence learning problem. Given an input sequence of scalars, the objective is to estimate the succeeding values of the sequence. An important task is to determine how many past values should be considered as the features under consideration, i.e., which slice dimension of the series allows for good model generalization. By nature, time series data is dynamic and subject to change over time, so an initial decision is not necessarily a sustainable solution. Designing ensembles of LSTM networks allows us to incorporate multiple sequence lengths into our prediction model. In the following, we explain how.

LSTMs with varying sequence lengths By architecture, LSTMs are only capable to process sequences of equal lengths per epoch, due to the required matrix operations in the optimization process. In many applications, however, varying sequence lengths are inevitable. One example is machine translation where the length of an input sentence can be arbitrarily long [21]. Padding is usually used to overcome that problem [12]. This implicitly means that, although two models trained with even slightly different sequence lengths have a large intersection of training data, they learn different yet related patterns. This constitutes a promising setting for ensemble learning.

Locating candidate sequence lengths In order to train a number of snapshot LSTMs with different sequence lengths, the first step is to identify the right choices of these. A naive approach is to select sequence lengths from a random distribution. To get sequence lengths that can catch effects of seasonality, we apply a fast fourier transformation (FFT) to the training data and estimate the power spectra [23]. The motivation behind this is that the FFT is an efficient

method to extract the right periodicities from a given time series. This allows the snapshots to encode different patterns, seasonalities, and other time-dependent effects in the series.

Generating a snapshot ensemble of LSTMs with varying sequences [13] conduct a variant of simulated annealing in order to adapt the learning rate and escape from local minima. In this case, a snapshot is a further optimization of its predecessor using the identical training data, which leads to a relatively low level of diversity across the snapshots. We propose another strategy in order to increase diversity: Instead of adapting the model parameters, we feed the LSTM with different slices of the data. This is possible because the dimensions of the training data must be identical within a single epoch but not for two separate epochs. Given a set $S = \{s_1, s_2, ..., s_n\}$ of different sequence lengths we store in total n snapshots of the LSTM. After each snapshot based on s_i, the training process is continued with a different data slice through time according to s_{i+1}. The final holdout estimates of the individual snapshots are commonly combined by taking the mean of the base forecasts. This assumes that each snapshot is equally important with respect to the combination of forecasts. In order to allow for more flexibility, we extend the mean function by a meta learner. Ridge Regression has proven to be an effective choice here [25]. The process of

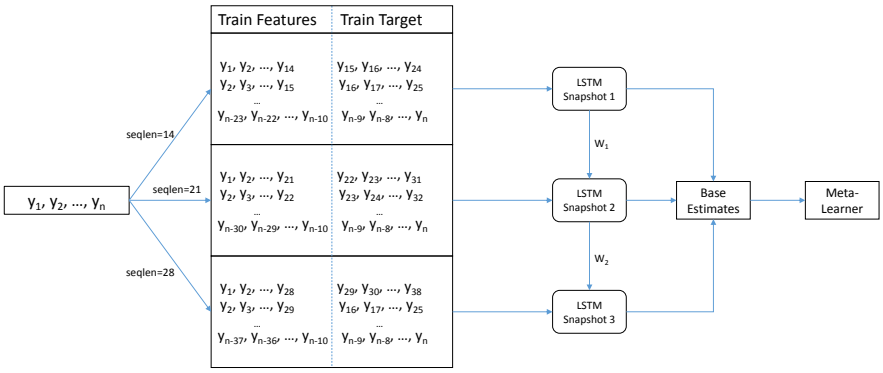


Fig. 1. LSTM Snapshot Training Framework

the ensemble construction at training time is depicted in Fig. 1 for the example case $S = \{14, 21, 28\}$ and a forecasting horizon of 10. First, the training data $y_1, ..., y_n$ (75% of the total data) is split according to the most potent sequence

lengths provided by the FFT (in decreasing order of FFT significance). In our experiments, we use the top 20 sequence lengths. Next, the first snapshot is trained with the respective data slices based on the first sequence length. We train each snapshot for five epochs and standardize the data by its z-transform prior to training. The base LSTM learners' architecture is set up of two LSTM layers with 64 and 128 neurons as well as 20% dropout. Adam is used as the optimizer with a learning rate of 0.001. The weight matrix of the first snapshot is then updated based on the data slices for the second sequence length, and so on. In total, training is done for $5 \cdot 20 = 100$ epochs. After all snapshots are trained, a ridge regression meta model learns how to combine the individual forecasts of the 20 snapshots. Analogously, at test time, all 20 base models provide their forecasts to the meta learner, which then combines them to the final estimate for the 10 step ahead forecasts.

4 Experiments

We test the proposed methodology on five data sets of different kind. We train a snapshot ensemble for each data set where we start with the strongest periodicity according to the FFT. Subsequently, each LSTM snapshot is based on the next strongest periodicity. In total, 20 snapshots are trained. An overview of the datasets is given in Table 1 and Fig. 2. Furthermore, Fig. 3 displays the power spectrum for the sunspots series. This example shows that there exist a number of unequally well suited periodicities. Each of these contains different patters which we aim to extract using snapshot ensembles. To show the effectiveness as well as the efficiency of our approach, the performance of the snapshot ensemble is measured against the following three baselines:

1. Independent LSTM ensemble. Instead of continuing the training process by escaping from a local minimum, the LSTM is reinitialized randomly and fed with the new data slices. Instead of n snapshots, we end up with n LSTMs whose training process was completely independent of one another. In contrast to this, a snapshot inherits its initial weights from its preceding snapshot.
2. Single optimized LSTM. The best sequence length according to the FFT is used for the optimization of a single LSTM over all epochs.
3. ARIMA with model selection based on the AIC.

Notably, the total number of epochs is identical for all the neural net approaches. Due to different slices of the training data, the total runtime of the latter approach can slightly differ from the ensemble methods in either direction.

4.1 Model Evaluation

We validate the performance of our approach on five different data sets listed in Table 1. Fig. 2 illustrates the series on their original scale. Evidently, each of the datasets has its very own characteristics and dynamics. While the daily birth

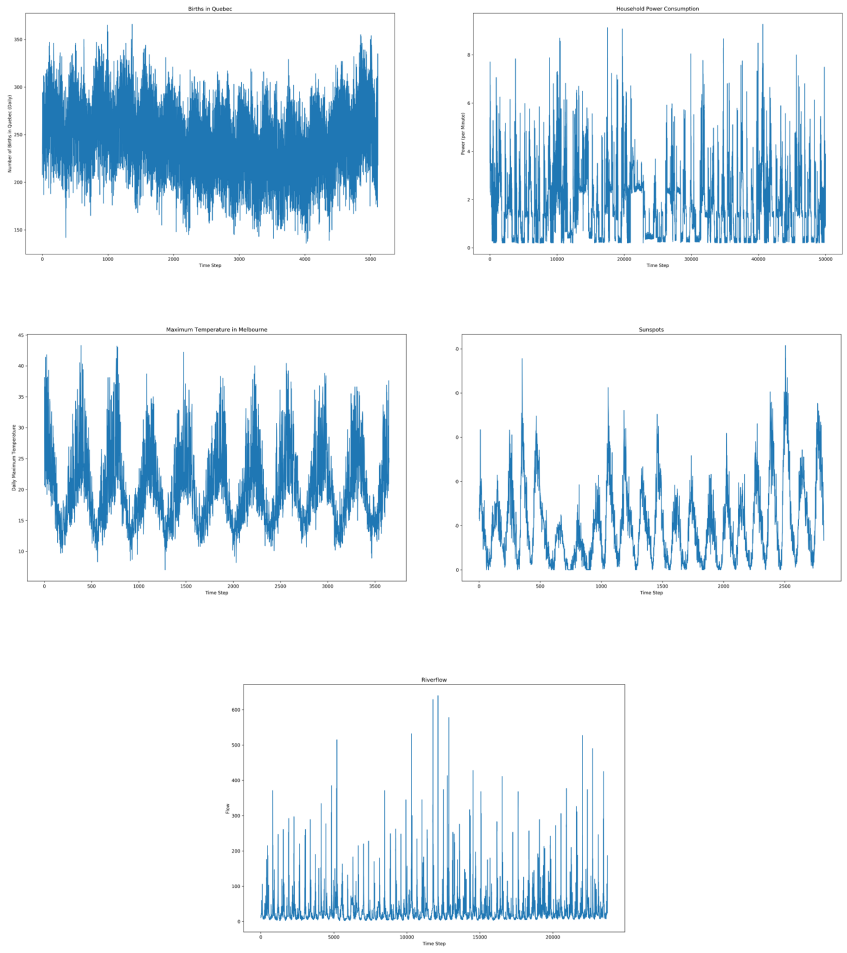


Fig. 2. Graphical Data Overview

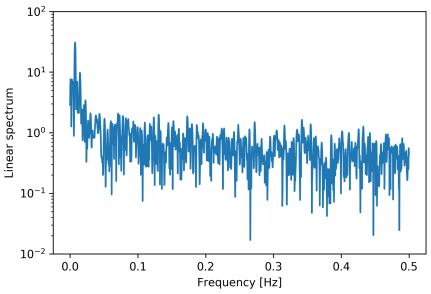


Fig. 3. Power Spectrum of the Sunspots Data Set

rates data set shows signs of weak stationarity, the sensor-generated household power dataset depicts more chaotic behavior with random noises. The latter is sampled by the minute. River flow, a monthly sampled time series, is clearly non stationary as well. The series of daily maximum temperatures repeats similar patterns over time as does the births data and shows clear signs of weak stationarity. Somewhere in between those cases fits the monthly sunspots data which shows seasonalities of varying strength and amplitude. Fig. 4 shows the root

Table 1. Datasets of the Experimental Analysis

Data	Number of Observations
Births in Quebec [9]	5,113
Household Power Consumption [16]	50,000
Maximum Temperature in Melbourne[1]	3,650
Number of Sunspots[1]	2,820
Riverflow[1]	23,741

mean square error (RMSE) on the holdout set of each dataset and method. Besides the performance of the stacked ensembles ('Snap Stack': stacked snapshot ensemble, 'ClassEns Stack': stacked ensemble of independently trained LSTMs), metrics for mean ensemble forecasts ('Snap Mean', 'ClassEns Mean') and single model forecasts ('Single opt.') are shown. The key outcomes of the analysis are:

- Snapshot ensembles with Ridge Regression as a meta learner outperform conservative ensembles as well as the single, optimized model in all cases.The traditional ARIMA models show inferior forecasting accuracy.
- On average, the stacked snapshot ensemble performs 4.2 % better than the next best baseline.
- The greatest performance gain obtained by the stacked ensemble is realized for the sunspots data. Here, the stacked snapshot ensembles outperform the next best method by 13.8%, while the performance win for the other four datasets is in a significantly lower range between 1.0% and 3.4%. Looking at the illustrated data in Fig. 2, this is an indication that our approach is particularly suitable for time series with seasonalities of varying intensity. Peaks of different amplitudes are handled well by the stacked snapshot ensemble, which a single model fails to do with a high degree of precision.
- Extending snapshot ensembles by the introduction of a meta learner leads to a great boost in performance compared to the simple mean combiner.
- The ensemble forecasts are significantly different from the estimates of the remaining models, based on the paired t-test for significance.
- The single optimized LSTM only shows comparative performance if the structure of the dataset is approximately stationary over time, as in the case of the maximum temperatures series. This supports our hypothesis that

[1] https://datamarket.com/data/list/?q=, accessed June 1, 2018

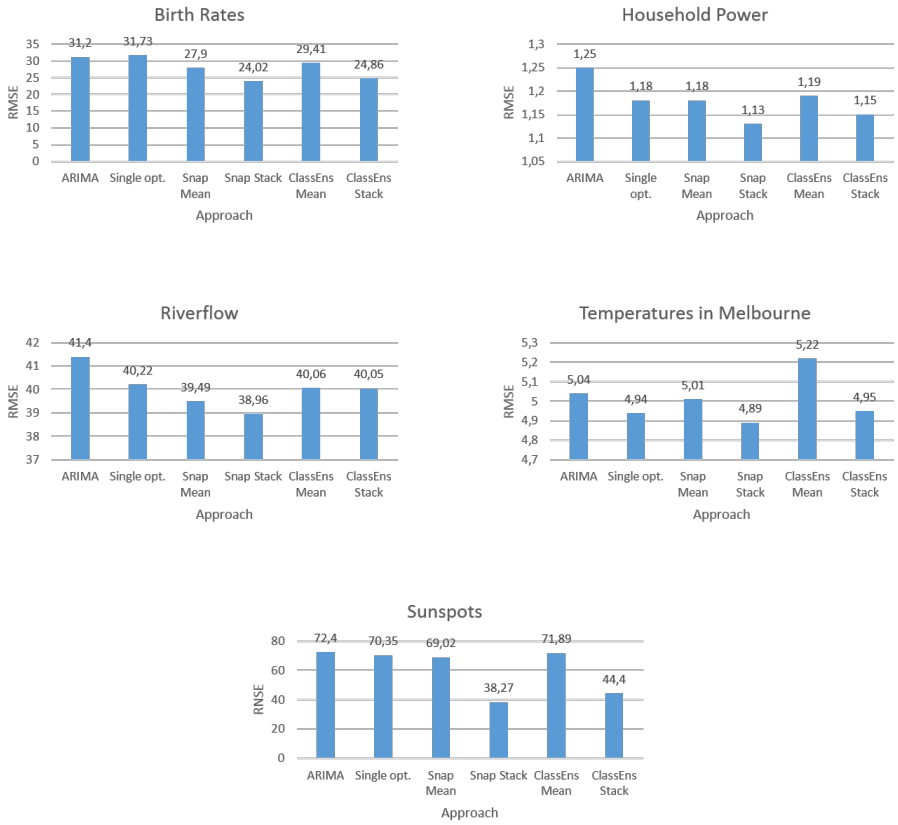


Fig. 4. Model Performance

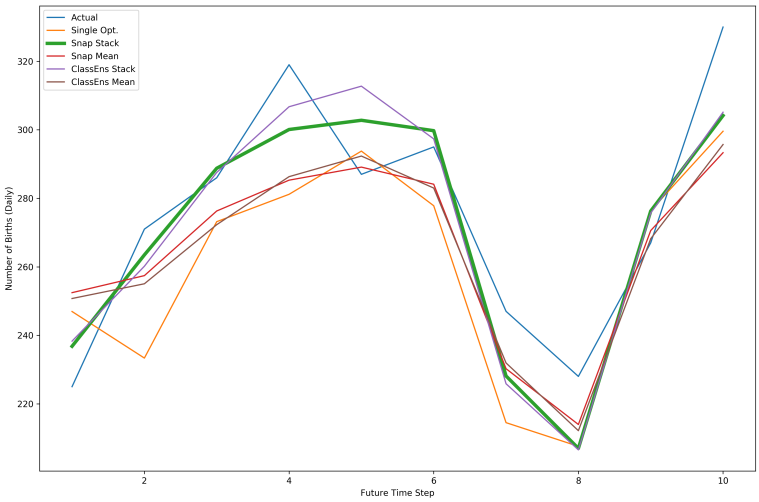


Fig. 5. Exemplary Forecast

snapshot ensembles are particularly suitable for cases where patterns are spread across multiple sequence lengths.
– Reslicing the input data according to the FFT after each snapshot leads to base learners with high diversity. This enables the meta learner to exploit different knowledge that is encoded across the snapshots. As an example, the ordered FFT sequence lengths for the birth rates dataset are as follows: 365, 183, 73, 61, 37, 91, 41, 30, 10, 52, 11, 26, 852, 28, 14, 568, 341, 16, 20, 465. This clearly shows how FFT extracts potent periodicities from the time series as the yearly and monthly seasonalities are immediately detected.

An exemplary 10 step ahead forecast is shown in Fig. 5. Here, the first holdout sequence of the birth rates series along with its model estimates is illustrated. One can see the significant improvements that are attributed to the meta learner, leading to the reduction in forecasting error.

The code for the experiments is available on GitHub[2] .

5 Future Work and Conclusion

Snapshot ensembles based on FFT sequence lengths are an efficient method to extract diverse patterns from data. We have shown that they yield superior forecasting performance in comparison to the standard optimization of a single

[2] https://github.com/saschakrs/TS-SnapshotEnsemble, accessed June 1, 2018

LSTM and an ensemble of fully independently trained LSTMs, without the need for additional computational costs. It turned out that these results are stable across different data sets, although the relative performance boost differs depending on the underlying data structure. Our approach enables the automated generation of robust time series forecasts without the assumption of a specified data distribution. This makes the framework a valuable application for systems that require the future estimation of one or more key performance indicators that develop over time.

There is further potential regarding the design of the ensemble architecture: Besides the configuration of the individual base learners, different combiner functions might improve the overall performance for certain problems. In addition to this, we found that five epochs per snapshot lead to good overall performance of the ensemble, however, this parameter could be higher for very complex learning tasks.

It is also possible to extend the ensemble by different model types. Integrating autoregressive models or state-space representations could increase model diversity and thereby lead to a greater performance win by the combiner function.

Finally, LSTM snapshot ensembles are currently limited to univariate time series. Evaluating their applicability to the multivariate case is another challenge worth investigating. It would also be interesting to evaluate the applicability of stacked snapshot ensembles to different sequence learning tasks such as machine translation.

References

1. Ratnadip Adhikari. 2015. A neural network based linear ensemble framework for time series forecasting. Neurocomputing 157 (2015), 231-242.
2. Ratnadip Adhikari and RK Agrawal. 2014. A linear hybrid methodology for improving accuracy of time series forecasting. Neural Computing and Applications 25, 2 (2014), 269-281.
3. Cagdas Hakan Aladag, Erol Egrioglu, and Cem Kadilar. 2009. Forecasting nonlinear time series with a hybrid methodology. Applied Mathematics Letters 22, 9 (2009), 1467-1470.
4. Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks 5, 2 (1994), 157-166.
5. Mohamed G Elfeky, Walid G Aref, and Ahmed K Elmagarmid. 2005. Periodicity detection in time series databases. IEEE Transactions on Knowledge and Data Engineering 17, 7 (2005), 875-887.
6. Felix A Gers, Douglas Eck, and Jrgen Schmidhuber. 2002. Applying LSTM to time series predictable through time-window approaches. In Neural Nets WIRN Vietri-01. Springer, 193-200.
7. James Douglas Hamilton. 1994. Time series analysis. Vol. 2. Princeton university press Princeton.
8. Zhen He, Shaobing Gao, Liang Xiao, Daxue Liu, Hangen He, and David Barber. 2017. Wider and Deeper, Cheaper and Faster: Tensorized LSTMs for Sequence Learning. In Advances in Neural Information Processing Systems. 1-11.

9. Keith W Hipel and A Ian McLeod. 1994. Time series modelling of water resources and environmental systems. Vol. 45. Elsevier.
10. Sepp Hochreiter and Jrgen Schmidhuber. 1997. Long short-term memory. Neural computation 9, 8 (1997), 1735-1780.
11. Kurt Hornik, Maxwell Stinchcombe, and Halbert White. 1989. Multilayer feedforward networks are universal approximators. Neural networks 2, 5 (1989), 359-366.
12. Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional neural network architectures for matching natural language sentences. In Advances in neural information processing systems. 2042-2050.
13. Gao Huang, Yixuan Li, Geoff Pleiss, Zhuang Li, John Hopcroft, and Kilian Weinberger. [n. d.]. Snapshot Ensembles: Train 1 Get M for Free. In Proceedings of the International Conference on Learning Representations (ICLR 2017).
14. Sascha Krstanovic and Heiko Paulheim. 2017. Ensembles of Recurrent Neural Networks for Robust Time Series Forecasting. In International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer, 34-46.
15. Martin Lngkvist, Lars Karlsson, and Amy Loutfi. 2014. A review of unsupervised feature learning and deep learning for time-series modeling. Pattern Recognition Letters 42 (2014), 11-24.
16. M. Lichman. 2013. UCI Machine Learning Repository. (2013). http://archive.ics.uci.edu/ml
17. Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. 2015. Long short term memory networks for anomaly detection in time series. In Proceedings. Presses universitaires de Louvain, 89.
18. Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. In International Conference on Machine Learning. 1310-1318.
19. Deepak Sharma, Biju Issac, GPS Raghava, and R Ramaswamy. 2004. Spectral Repeat Finder (SRF): identification of repetitive sequences using Fourier transformation. Bioinformatics 20, 9 (2004), 1405-1412.
20. R.H. Shumway and D.S. Stoffer. 2010. Time Series Analysis and Its Applications: With R Examples. Springer New York.
21. Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems. 3104-3112.
22. LiWang, Haofei Zou, Jia Su, Ling Li, and Sohail Chaudhry. 2013. An ARIMA-ANN hybrid model for time series forecasting. Systems Research and Behavioral Science 30, 3 (2013), 244-259.
23. Peter Welch. 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on audio and electroacoustics 15, 2 (1967), 70-73.
24. G Peter Zhang. 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50 (2003), 159-175.
25. Le Zhang and Ponnuthurai Nagaratnam Suganthan. 2017. Benchmarking Ensemble Classifiers with Novel Co-Trained Kernal Ridge Regression and Random Vector Functional Link Ensembles [Research Frontier]. IEEE Computational Intelligence Magazine 12, 4 (2017), 61-72.
26. Mariana Oliveira and Luis Torgo. 2014. Ensembles for time series forecasting. JMLR: Workshop and Conference Proceedings 39:360-370
27. Vitor Cerqueira, et al. 2017. Arbitrated Ensemble for Time Series Forecasting. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham.

Implications for Aggregate Inflation of Sectoral Asymmetries: an Empirical Application

Hannu Koskinen [11] and Jouko Vilmunen [2]

[1] University of Tampere, Faculty of Management, 33014 University of Tampere, Finland
hannu.koskinen@uta.fi
[2] University of Turku, Turku School of Economics, Finland
jouko.vilmunen@utu.fi

Abstract In order for the forecast and policy analyses concerning the business cycle fluctuations of the economy to be precise, the sector specific characters together with the steady state of the economy have to be estimated correctly. This steady state, i.e. the potential output of the economy, is determined by the preference parameters of the agents. Following the theoretical identification scheme for a two-sector DSGE macro model presented in Woodford [10] and further modified in a study by Koskinen and Vilmunen [5] the focus of this study is to estimate and identify the possibly divergent sector specific parameters which are of interest. A problem is, that the estimation procedure of these parameters is conditional for at yet unknown steady state. Therefore we propose and utilize a two-step estimation procedure for the parameters of interest. In this modelling approach we analyse and compare two different sectors of the Finnish economy, manufacturing industry and building industry, each in turn to the rest of the economy during 2000:Q1- 2015Q2. As the relevant parameters within a sector reflect the divergent preferences of the economic agents to the goods produced at different sectors, the relative price movements and adjustment asymmetries stemming from these kind of divergences has a central role in allocational efficiency and welfare of the economy. As a result we find out the importance of the relative price movements together with group specific preference parameters for the adjustment of the economy after hit by a shock.

Keywords: A Two Sector DSGE model, Bayes-estimation, Sector Specific Parameters.

1 Introduction

From the stabilization policy point of view, it is almost a necessity to understand the different characters of the particular sectors within the economy. These characters could be understood to express the nature of the preferences of the representative agent in the economy. As the preferences to the goods produced as well as market mechanism in general could diverge between the sectors, there is potential for sectoral adjustment

[1] Koskinen gratefully acknowledges financial support from the Finish Cultural Foundation and wish to thank Kari Heimonen and Maritta Paloviita for insightful comments.

asymmetries in an economy after hit by a shock. This is due to a fact that the adjustment to any shock that hits the economy reflect, in turn, the structural factors of that economy. In addition, as such, the economy could been characterised to be compounded of separate sectors, each of which having its own structural features. In an economy with nominal price stickiness where the frequency of individual price changes as well as price elasticities (and then the mark-ups) are sector specific, inflation distorts relative prices. This cause's economic inefficiency in terms of a loss of consumer welfare as output fluctuates around its natural level. Then, as most of the economists believe, the policy makers should respond to these deviations by actively dampening them. However, the response of the economy (output/inflation) to e.g. monetary policy shock reflect to some extent the structural factors that characterize the economy. These structural features could diverge between particular sectors in that economy, and hence the response is sector specific.

This study, then, estimates those sector specific parameters that are crucial for inflation dynamics in the economy studied. This, in turn, enables us to analyse the aggregate as well as sector specific dynamics of the economy as it is hit by various structural shocks. Although we are dealing with a model for the business cycle fluctuations, estimation of the long-term trend (i.e. the steady state / potential output) has a crucial role here. This is due to a fact that this trend determines the phase of the business cycle and, moreover, the estimation of the parameters of interest is conditional to the steady state which, in turn, is determined by these parameters. Therefore we proceed in a two-step manner in our Bayesian estimation procedure; the first estimation round, with calibrated parameter values as starting point, serves as a starting point for a second round. The main interest is on the elasticity of substitution (which relates to the magnitude of a relative cost-push shock) and autoregressive coefficient (which determines the duration of a shock) parameters. Shocks that are for our interest are a general cost-push and an interest rate shocks. These shocks are seen to be stemming from the preferences of the representative household (agent) as the price setting behaviour of the agent determines the magnitude of these shocks, and moreover, the price setting is subject to the elasticity of substitution parameter.

Although the idea of a multi- or two sector model for aggregate macro modelling is a familiar one from several earlier macroeconomic studies, see e.g. Woodford [10] and Tille [8] for examples of the approach, here the modelling framework is an extended version of a dynamic stochastic general equilibrium (DSGE) model. In this framework, we allow for the underlying preference parameters to diverge between the sectors studied. Earlier studies of the inflation persistence and dynamics have concentrated to study the role of the price rigidity, allowing it to differ across the sectors studied, see e.g. Bouakez, Cardia and Ruge-Murcia [2]. Hence, this study will offer a further and deeper understanding of how the disclosed micro-level heterogeneity should affect macro-level analysis.

This paper is a shortened version of the study by Koskinen and Vilmunen [6] where the full derivations of the model are represented. Model estimations indicate that we can distinguish those sector specific parameters and output trend levels (steady states) that are crucial for policy analyse.

2. **A Two Sector Model**

As in the previous study of Koskinen and Vilmunen [5], the economy is composed of two sectors within which the goods are imperfectly substitutable. Hence, there is imperfect competition in the relevant markets. A representative household in this economy derives utility from a consumption bundle that is a CES aggregate of the sector specific consumption indices. These sector specific consumption indices are CES aggregates over a continuum of individual goods. Our representative household also works, thus generating disutility in the usual manner. The (flow) budget constraint determines the feasible choices for our representative household: on top of allocating income on consumption, the household can invest in one period bonds, which generates interest income. By working, the household earn wage income. We thus assume that the representative household seeks to maximise the following intertemporal utility

$$E_0 \sum_{t=0}^{\infty} \beta^t U_t \tag{1}$$

where β is the discount factor and where

$$U_t = \varphi_t^B \left(\frac{1}{1-\sigma} (C_t - h_t C_{t-1})^{1-\sigma} - \frac{1}{1+\psi} (L_t)^{1+\psi} \right) \tag{2}$$

σ is the inverse of the inter-temporal elasticity of substitution in consumption and ψ the inverse of the Frisch elasticity. *Ct* is now an index of the household's consumption of the goods that are supplied, while L_t is the labor supply. Eq (2) contains also a general preference shocks φ_t^B. The external habit formation is captured by the term *h*.

Households maximise their objective function (1) subject to the (flow) budget constraint, where total nominal income consists of two components: labour income ($P_t W_t L_t$) and the gross return on the bonds ($R_t^b B_t$). Then P_t is the price level and B_t denotes bonds.

As the capital stock is assumed to be fixed in the considerations we do not include it here. This is because we will focus on the dynamic effects of cost push and interest rate shocks at the business cycle frequency. The underlying assumption here is that variations in the capital stock are not the main driver for business cycles.

Consumption behavior

The Euler condition for the optimal intertemporal allocation of consumption is derived from the maximization problem of the objective function (2) subject to budget constraint with respect to consumption and (nominal) return on bonds R_t^b. This yields to the usual consumption Euler equation describing the marginal utility of consumption.

Then we end up with the optimal consumption dynamics of the log-linear form:

$$\hat{c}_t = \frac{h}{1+h} \hat{c}_{t-1} + \frac{1}{1+h} E_t \hat{c}_{t+1} - \frac{1-h}{\sigma\ (1+h)} [\hat{r}_t - E_t \hat{\pi}_{t+1}] + \hat{\varphi}_t^B . \tag{3}$$

The hatted variables represent log deviations from the steady state.

2.1 Deriving a model for sectoral asymmetries

The aggregate consumption index C_t is of the CES-form and consists of two sub-indexes for the commodity groups n_1 and n_2 as described earlier in Koskinen and Vilmunen [5];

$$C_t = \left[(n_{1t}\varphi_t^{n1})^{\frac{1}{\eta}} C_{1t}^{\frac{\eta-1}{\eta}} + (n_{2t}\varphi_t^{n2})^{\frac{1}{\eta}} C_{2t}^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}} \tag{4}$$

where $\varphi_{j,t}$ is a shock to the relative weight of the commodity group in a consumption basket and η is the elasticity of substitution between the groups and the sectoral consumption index aggregates a continuum of sector-specific goods

$$C_{jt} = \left[n^{\frac{-1}{\theta_j}} \int_0^n c_t(i)^{\frac{\theta_j - 1}{\theta_j}} di \right]^{\frac{\theta_j}{\theta_j - 1}} \qquad j=1,\ 2 \quad N_1 = [0,\ n1\ ],\ N_2 =]n1\ ,\ 1] \tag{5}$$

Here θj is the elasticity of substitution between sector j goods defining the own price elasticity of the demand for these goods. We allow the two θ:s to differ. Sectoral price indices, which defines the minimum cost of buying a unit of the sector j good, satisfy

$$P_{j,t} = \left[n_{jt}^{\frac{-1}{\theta_j}} \int_{N_j} p_{j,t}(i)^{1-\theta_j} di \right]^{\frac{1}{1-\theta_j}}, \ j = 1,2 \tag{6}$$

whereas the aggregate price index corresponding to the aggregator in (6) is given by

$$P_t = \left[n_{1t}\varphi_t^{n1} P_{1,t}^{1-\eta} + n_{2t}\varphi_t^{n2} P_{2,t}^{1-\eta} \right]^{\frac{1}{1-\eta}} \tag{7}$$

Optimal allocation for different goods in sector j = 1, 2 can be derived from the minimization problem

$$\min \int p_{j,t}(i) c_{j,t}(i) di \quad \text{s.t.} \quad \left[n_j^{\frac{-1}{\theta_j}} \int_{N_j} c_{j,t}(i)^{\frac{\theta_j - 1}{\theta_j}} di \right]^{\frac{\theta_j}{\theta_j - 1}} \geq C_{jt}$$

and its first order conditions. Demand for different brands c(i) within a group j is then

$$c_{jt}(i) = n_{jt}^{-1} \left[\frac{p_{jt}(i)}{P_{jt}} \right]^{-\theta_j} C_{jt}. \tag{8}$$

The sectoral market demand functions in equilibrium are then

$$C_{jt} = n_{jt}\varphi_t^{nj} \left(\frac{P_{jt}}{P_t} \right)^{-\eta} C_t \tag{9}$$

As in the previous study the aggregators have been normalized so that under the common individual price in both sectors,

$$p_{jt}(i) = p_t \forall j,i,\ c_{jt}(i) = \varphi_{jt} C_t$$

Disutility of labour is given by, $v(h(i); \xi_{jt})$ where ξ is a vector of parameters of interest, so that sector-specific shocks to preferences regarding labour supply is allowed for, but not good-specific. Production of the good *i* is obtained via the production function

$$y_{jt}(i) = A_{jt} f(h_t(i)) \tag{10}$$

which thus implies that sector-specific shocks are allowed for. We assume wage-taking firms and that the (nominal) profits of firm *i* in sector *j* can be written as in the previous study together with the real total cost and real marginal cost, s_t^j, of supplying any good *i* in sector *j* can be represented as earlier.

Note that we have used the assumption that the households are on their labour supply schedule so that the real wage equals the marginal rate of substitution between labour and consumption. The desired mark-up in sector *j* is now $\mu_j = \frac{\sigma}{\sigma - 1}$, a constant, but sector specific. The natural level of output in sector *j*, $Y_{j,t}^n$, is defined as the common level of sector *j* output under flexible prices[2]. It satisfies

$$\mu_j s^j\left(Y_{jt}^n, Y_t^n; \tilde{\xi}_t\right) = \frac{P_{jt}}{P_t}\left(\frac{Y_{jt}^n}{n_j \varphi_{jt} Y_t^n}\right)^{-\frac{1}{\eta}} \tag{11}$$

with the intended interpretation that the utmost right hand side indicates the relative price $P_{j,t} / P_t$ required to induce the relative demand $Y_{j,t}^n / Y_t^n$. The natural rate of aggregate output Y_t^n aggregates sectoral natural outputs according to the CES-aggregator. If $\tilde{\xi} = 0$ and φ =1 for all *t* and for both sectors, the flex price equilibrium involves a common output $\bar{Y}$ for all goods (satisfying the above equilibrium pricing equation)

$$s^j(\bar{Y}, \bar{Y}; 0) = \frac{1}{\mu_j}\left(\frac{1}{n_j}\right)^{-\frac{1}{\eta}}.$$

Log-linearizing around this equilibrium gives us the optimality conditions for the real marginal cost as in our previous study.

Assume Calvo-type (Calvo [1]) price staggering in each of the two sectors with α_j the fraction of goods prices that remain constant in any given period in sector *j*. A firm *i* in that sector that is lucky to get the change to optimize its price in period *t* chooses its new price $p_t(i)$ to maximize the expected present value of its profits

$$E_t\left\{\sum_{T=t}^{\infty} \alpha_j^{T-t} Q_{t,T}\left[\prod_T^{ji}(p_t(i))\right]\right\} \tag{12}$$

The F.O.C for this programme is, after log-linearizing, given by

[2] For derivation, see e.g. Walsh [9] Ch. 8.2.

$$E_t \sum_{T=t}^{\infty} (a_j \beta)^{T-t} \left\{ \hat{p}^*_{jt} - \left[\hat{s}^j_{t,T} - \hat{p}_{jT} + \sum_{\tau=t+1}^{T} \pi_{j\tau} \right] \right\} = 0 \tag{13}$$

where $\hat{p}^*_{j,t} = \ln\left(\frac{P^*_{j,t}}{P_{j,t}}\right)$ denotes the relative price (relative to others in sector *j*) of the firms that get to optimize their price at date *t* and $\hat{p}_{j,t} = \ln\left(\frac{P_{j,t}}{P_t}\right)$ is the real price of sector *j* at date *t* (i.e. the relative date *t* price of sector *j* relative to "cpi" or overall price level). On the other hand, $\hat{s}^j_{t,T}$ is the real marginal cost of the firms in sector *j* that last change their prices at date *t*. Noticing the decomposition of this real marginal cost together with the expression for the sectoral price indexes in eq. (6) a log-linear approximation (around the steady state) allows us to derive the following relationship

$$\begin{aligned} 0 &= \left[(1-\alpha_j)\hat{p}^*_{jt} - \alpha_j \hat{\pi}_{jt}\right](1-\theta_j) \Leftrightarrow \\ \hat{p}^*_{jt} &= \left(\frac{\alpha_j}{1-\alpha_j}\right)\hat{\pi}_{jt} \end{aligned} \tag{14}$$

Now, insert everything into the optimal pricing eq. (11) to give us

$$\begin{aligned} \hat{\pi}_{jt} &= \underbrace{\left(\frac{(1-\alpha_j\beta)(1-\alpha_j)}{\alpha_j(1+\omega\theta_j)}\right)}_{\zeta_j}\left(\hat{s}^j_t - \hat{p}_{jt}\right) + \beta E_t \hat{\pi}_{j,t+1} \\ &= \zeta_j(\hat{s}^j_t - \hat{p}_{jt}) + \beta E_t \hat{\pi}_{j,t+1}, \end{aligned} \tag{15}$$

where ω denotes the elasticity of the real marginal cost function with respect to $\hat{y}_{jt}$ (and $\hat{y}^n_{jt}$). We are almost there, as we still need to express the sector specific real marginal cost in terms of the relevant average sectoral output measure. From the expression for the demand for the sectoral composite good we obtain

$$\begin{aligned} \hat{y}_{jt} &= \hat{\varphi}^{nj}_t + \hat{y}_t - \eta\hat{p}_{jt} \\ \hat{y}^n_{jt} &= \hat{\varphi}^{nj}_t + \hat{y}^n_t - \eta\hat{p}^n_{jt} \end{aligned} \tag{16}$$

where $\hat{y}_t \equiv \log(Y_t / \bar{Y})$ and $\hat{y}^n_t \equiv \log(Y^n_t / \bar{Y})$ are log deviations from the steady state. Substituting this into the average real marginal cost of sector j gives us

$$\hat{s}^j_t - \hat{p}_{jt} = (\sigma^{-1} + \omega)(\hat{y}_t - \hat{y}^n_t) - (1-\omega\eta)(\hat{p}_{jt} - \hat{p}^n_{jt}) \tag{17}$$

Inserting this into (15), we obtain

$$\hat{\pi}_{jt} = \zeta_j(\sigma^{-1} + \omega)(\hat{y}_t - \hat{y}_t^n) + \gamma_j(\hat{p}_{Rt} - \hat{p}_{Rt}^n) + \beta E_t \hat{\pi}_{j,t+1}, \tag{18}$$

where $\hat{p}_{Rt}^n \equiv \frac{1}{\eta}\left[\left(\hat{\varphi}_t^{n2} - \hat{\varphi}_t^{n1}\right) - \left(\hat{y}_{2t}^n - \hat{y}_{1t}^n\right)\right]$ and the sectoral relative price

$$\hat{p}_{Rt} = \ln\left(\frac{P_{2t}}{P_{1t}}\right) = \ln\left(\frac{P_{2t}}{P_{2,t-1}} \cdot \frac{P_{1t-1}}{P_{1,t}} \cdot \frac{P_{2,t-1}}{P_{1,t-1}}\right) = \hat{\pi}_{2t} - \hat{\pi}_{1t} + \hat{p}_{R,t-1} \tag{19}$$

is obtained from the aggregate price index, note that we can represent the sectoral inflation differential in terms of the lagged relative price $\hat{p}_{R,t-1}$. Now define $\kappa_j = \zeta_j(\sigma^{-1} + \omega)$ and since $\gamma_1 = n_2\zeta_1(1+\omega\eta)$ and $\gamma_2 = -n_1\zeta_2(1+\omega\eta)$ we can write

$$\hat{\pi}_{jt} = \kappa_j(\hat{y}_t - \hat{y}_t^n) + \gamma_j(\hat{p}_{Rt} - \hat{p}_{Rt}^n) + \beta E_t \hat{\pi}_{j,t+1} \tag{20}$$

Aggregate inflation rate $\hat{\pi}_t = n_1\hat{\pi}_{1t} + n_2\hat{\pi}_{2t}$ has a similar representation

$$\hat{\pi} = \bar{\kappa}(\hat{y}_t - \hat{y}_t^n) + \bar{\gamma}(\hat{p}_{Rt} - \hat{p}_{Rt}^n) + \beta E_t \hat{\pi}_{t+1} \tag{21}$$

$$\bar{\kappa} = n_1\kappa_1 + n_2\kappa_2, \quad \bar{\gamma} = n_1\gamma_1 + n_2\gamma_2,$$

where we can use equation (19) for the period *t* relative price to ascertain that aggregate inflation in general depends on the lagged relative price and not only on aggregate variables. Note also that identical sectoral price adjustment frequency – $\alpha_1 = \alpha_2$ – is not anymore sufficient to eliminate the dependence of aggregate inflation on the sectoral relative price. On the other hand, if both the frequency of price adjustment and price elasticity of demand are equal across sectors, then we have $\bar{\gamma} = n_1\gamma_1 + n_2\gamma_2 = 0$ so that aggregate inflation is independent of the sectoral relative price.

From now on we assume, that output is given by a linear production function and it consist of goods to be consumed so that *Y*= *C* because in the setting below it does not make a difference to distinguish production to consumption and investment goods, we further assume that capital is firm specific and fixed at the business cycle fluctuations at least, see Woodford (2005). Hence

$$C_t = A_t L_t,$$

where we assume, that technical progress and productivity are catch by

$\frac{A_t}{A_{t-1}} \equiv dA_t = \exp z_t$, where z_t is an AR(1) process of the form $z_t = \alpha_t + \rho z_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$.

Here the natural, or potential, level of output (the level with flexible prices) is given by $c_t^n = \log\left((\frac{\theta}{\theta-1})^{-1/\sigma+\psi} A_t^{\frac{1+\psi}{\sigma+\psi}} \right)$ in the symmetric case, for derivation see e.g. Walsh. In a case with asymmetries (with different thetas θ_j, j= 1, 2) we have:

$$\hat{c}_t^n = \lambda_{A1}\hat{a}_{1t} + \lambda_{A2}\hat{a}_{2t} + \lambda_{\varphi 1}\hat{\varphi}_t^{nj} + \lambda_{\varphi 2}\hat{\varphi}_t^{nj},$$

$$\text{where } \lambda_{Aj} = \frac{n_j^{1/\eta}(\mu_j)^{\frac{1-\eta}{\psi\eta}}}{n_1^{1/\eta}(\mu_1)^{\frac{1-\eta}{\psi\eta}} + n_2^{1/\eta}(\mu_2)^{\frac{1-\eta}{\psi\eta}}} \frac{(1+\psi)}{(\sigma+\psi)}$$

$$\text{and } \lambda_{\varphi j} = \frac{n_j^{1/\eta}(\mu_j)^{\frac{1-\eta}{\psi\eta}}}{n_1^{1/\eta}(\mu_1)^{\frac{1-\eta}{\psi\eta}} + n_2^{1/\eta}(\mu_2)^{\frac{1-\eta}{\psi\eta}}} \frac{\psi}{(\eta-1)(\sigma+\psi)}$$

for j = 1, 2. Here the productivity shocks $\hat{a}_{j,t}$ and the shocks to the relative weights of the sectors $\hat{\varphi}_t^{nj}$ causes fluctuations around the steady state. We notice, that (the inverse of) the inter-temporal elasticity of substitution in consumption together with (the inverse of) the Frisch elasticity of labour supply and the sector specific mark-up factors μ_j:s plays a crucial role in determining the level of the potential output. For derivation of these equations, see Koskinen and Vilmunen [6].

3. Data and estimation

Here we briefly introduce the estimation methodology used together with its theory and DSGE model solution strategy when implementing the estimation. Then we introduce the data used together with priors of the parameters for the estimation procedure.

3.1Bayesian estimation

The linearized system of solved DSGE model described above is estimated by Bayesian method. As this estimation methodology is a primary tool nowadays in macroeconomics, we introduce only briefly the necessary steps involved in our estimation procedure. Estimation in this case is based on the likelihood generated by the DSGE system of equations of observed variables. The primary goal of this study is to emphasize the possibility of the underlying model to estimate and distinguish the sector specific parameters. These parameters of interest will characterize the potential differences of the

sector compared to rest of the economy. Here we do not emphasize the shock and variance decomposition behaviour of the variables, these would be an important topic in the future research of this model of course. Therefore, in this study, the shocks (and measurement errors as we didn't add the full set of shocks) are added to avoid stochastic singularity.

Priors and the observed data are used to form a density function of the parameters of the model and the likelihood function which describes the density of the observed data. The priors are a priori beliefs of the weights of the parameters on the likelihood function. Priors and likelihood function are combined together by Bayes rule to form the posterior density function. This posterior density function which describes all the posterior moments of interest, is estimated and formed with the help of Kalman filter and Metropolis-Hastings sampling-like method. To initialize the Metropolis-Hastings algorithm to obtain the posterior distribution of our parameters we utilize a Monte-Carlo based optimization routine posterior mode as a starting point.

Finally, after deriving the posterior distribution of our particular model, we may have several (with the same data set) estimated competing models. The Bayes Factor provides a particularly natural method of compare and test these models against each other given the same data set (one model does not have to be nested within the other). As in our case, where we estimated some slightly different specifications of our model, we simply compare the ratio of posterior odds.

3.2. Data and priors

The empirical estimation of the parameters of interest is done using Finnish macroeconomic seasonally adjusted quarterly data over the period 2000:1 to 2015:2. The data are drawn from the Statistics Finland data base on (aggregate) gross domestic production, producer price index and the total number of employees. For the data on building industry and manufacturing industry we have the same sectoral variables. The data for interest rate is 3 month money market interest rate (Euribor). All the variables in the analysis and listed below are constructed from these as follows: observation data series for the output gap $c_t^{obs} = \widetilde{c}_t$ are demeaned log first differences of real per capita output series

$$c_t^{obs} = \widetilde{c}_t = \hat{c}_t - \hat{c}_t^n = (c_t - c_t^n) - \bar{c} \,,$$

and the observation data series for inflation $\pi_t^{obs} = \widetilde{\pi}_t$ and $\pi_{j,t}^{obs} = \widetilde{\pi}_{j,t}$ are demeaned log first differences of respective price level indexes etc.

In order to get the parameters defining the natural level of output, we estimated the full set of the model in two steps. First we took a prior for the output gap measure and estimated the parameters conditional to that prior value. Then, after deriving the full set of the parameters which could be identified within this model specification, we counted the new output gap measure and then, in the second step, estimated the parameters conditional to that new value.

Demeaning of all the other variables data gives us the observation time series used, for details see e.g. Pfeifer [7].

What is the role for demeaning? When estimating the model, in order to avoid a situation that the shocks are forced to account for a positive mean in the observed series these series have to be demeaned. In other words, we match an observed variable (which have mean 0) to a model variable which also has mean 0 (because it is a deviation from steady state).

The focus of this paper is on aggregate and sectoral inflation adjustment block of the economy π , π_j where we have aggregate and sectoral output c , c_j and the market rate of interest as other variables. Then we have the following linearized system of equations for the inflation adjustment:

$$\tilde{\pi}_{j,t} = \kappa_j(\tilde{c}_t) + \gamma_j(\tilde{p}_{R,t}) + \beta E_t \tilde{\pi}_{j,t+1} + \varphi_{j,t}^{pj} \quad , \quad \tilde{\pi}_t = n_1 \varphi_t^{nj} \tilde{\pi}_{1,t} + n_2 \varphi_t^{nj} \tilde{\pi}_{2.t}$$

$$\tilde{c}_{j,t} = \frac{h_j}{1+h_j}\tilde{c}_{j,t-1} + \frac{1}{1+h_j}E_t\tilde{c}_{j,t+1} - \frac{1-h_j}{\sigma\,(1+h)}[\tilde{r}_t^b - E_t\tilde{\pi}_{j,t+1}] \ , \ \tilde{c}_t = \tilde{c}_{1,t} + \tilde{c}_{2,t}$$

and

$$\tilde{c}_{j,t} = n_j\left(-\eta\left(\tilde{p}_t^{dj}\right) + \tilde{c}_t\right),$$

where $\tilde{p}_t^{dj} = \left[\ln\left(\frac{P_{j,t}}{P_t}\right) - \ln\left(\frac{P_{j,t-1}}{P_{t-1}}\right)\right] - \bar{p}$ and $\tilde{c}_t = \tilde{c}_{1,t} + \tilde{c}_{2,t} + \varphi_t^{dj}$. The demeaned market rate of interest evolves as $\tilde{r}_t^b = \tilde{r}_{t-1}^b + \varphi_t^r$.

There we have four kinds of exogenous shocks: $\varphi^{,pj}$ a general cost push shock, $\varphi^{,nj}$ a shock to relative weights, φ^r an interest rate shock to bond rate and $\varphi^{,dj}$ a demand shock. These shock variables are assumed to follow an independent first-order autoregressive stochastic process, e.g. $\varphi_t^{pj} = \rho\varphi_{t-1}^{pj} + u_t^{pj}$,

and the standard errors of the shocks are derived from the demeaned data series. Here we do not made any assumptions about the covariance of the shocks.

Regarding the choice of calibrated parameters (and the prior distributions, not shown here) we made the following choices which ably in every sectoral comparison (i.e. we compare manufacturing industry to the rest of the economy and building industry to the rest of the economy each in turn). The model specification is the same in every case. The calibrated values are for variables: β-discount factor (0.99), ρ – autoregressive coefficient for exogenous shocks (0.85), n_i – the relative steady-state share of manufacturing industry (0.2246), n_b – the relative steady-state share of building (0.068), Calvo parameter α = 0.75, intertemporal elasticity of substitution σ = 1.2, ω = 1, θ = 6 for within category *a* goods (the sector we compare to the rest of the economy), θ = 15 for within category *b* goods (the rest of the economy) and ψ = 2 the inverse of the Frisch-elasticity. Those values and the priors (for details, see Koskinen and Vilmunen [6]) are derived from earlier studies of Koskinen [4] and Kilponen and Ripatti [3].

More over the priors for the weighting values of parameters kappa (κ) and gamma (γ) in inflation adjustment equation have been set according to these parameter values.

4. Empirical results

As mentioned, we derived parameters for the output gap measure at the first step of estimation procedure and then evaluated this gap in the case of manufacturing industry to be $\bar{c}_{Ind}^{n} = 1{,}054$, this is the ratio of the natural level of output to the realised output level. Then we estimated the model parameters conditional to the output gap defined by this level of natural output.

Table 1. Estimated parameters for Industry

parameter	posterior distribution: mode	std. dev (Hessian)	mean	10%	90%	
habit (h_a)	0.7645	0.1240	0.7407	0.5366	0.9388	
habit (h_b)	0.8233	0.0918	0.8117	0.6658	0.9573	
weight (κ_a)	0.0179	0.0089	0.0224	0.0091	0.0368	
weight (κ_b)	0.0062	0.0047	0.0081	0.0015	0.0143	
weight (γ_a)	0.0916	0.0795	0.1414	0.0141	0.2580	
weight (γ_b)	0.0558	0.0198	0.0625	0.0309	0.0960	
Inter temp. elast. subst. (σ)	1.2415	0.3209	1.3212	0.7858	1.8410	
el. of subst. between groups (η)	5.7032	0.5612	5.7510	4.8205	6.6279	
auregr. coeff. (μ_a)	0.8409	0.0988	0.7896	0.6309	0.9611	
auregr. coeff. (μ_b)	0.6402	0.1135	0.6239	0.4344	0.8193	

In case of building industry the natural level of output is $\bar{c}_{build}^{n} = 1{,}029$ so it is approximately 3% higher than the existing level of output.

Table 2. Estimated parameters for Building

parameter	posterior distribution: mode	std. dev (Hessian)	mean	10%	90%	

habit (h_a)	0.8554	0.0740	0.8406	0.7336	0.9611	
habit (h_b)	0.7677	0.1230	0.7301	0.5408	0.9469	
weight (κ_a)	0.0185	0.0091	0.0219	0.0075	0.0348	
weight (κ_b)	0.0082	0.0028	0.0090	0.0042	0.0131	
weight (γ_a)	0.0871	0.07	0.1328	0.0157	0.2468	
weight (γ_b)	0.0525	0.0206	0.0602	0.0281	0.0914	
Inter temp. elast. subst. (σ)	1.2681	0.3175	1.3094	0.7774	1.8398	
el. of subst. between group (η)	6.4756	0.7287	6.6932	5.4510	7.8638	
auregr. coeff. (μ_a)	0.5343	0.1019	0.5354	0.3802	0.6941	
auregr. coeff. (μ_b)	0.8460	0.1046	0.7916	0.6280	0.9460	

We also slightly modified these models by incorporating different options for the Calvo parameters and we proceed as follows.

As mentioned earlier a natural way to evaluate the (prediction) performance of these models over each other is by Bayes Factor. In our case this factor is in every case approximately 1 (1 +- 0,0001) as the likelihood and the log data densities are nearly the same in every model variation. Then we can't distinguish which of the specifications with different Calvo parameter is better in the sense of (backward looking) weighting parameters kappa (κ) and gamma (γ). This may be a consequence of the forward looking nature of our inflation adjustment model specification, which put more weight on the expected inflation. Hence we do not report the outcomes with these different parameter values as they seems to give no additional information about the 'true' model. However, our inflation adjustment model seems to perform well as the one-step ahead linear Kalman filter forecast figures shows.

The diagnostic figures (in case of Indusry) for convergence of the parameters, goodness of fit (smoothed and historical values of the variables) by one-step ahead linear Kalman filter forecast evaluated at the posterior mode together with smoothed one-step ahead forecast of shock variables are presented upon request and in Koskinen and Vilmunen [6].

5. Conclusions

As relative price movements has a crucial role in resource allocation and market mechanism in general, the performance of stabilization policy is not invariant on the underlying parameters that determine the structure of this mechanism. As the parameters and hence the structure could diverse between the sectors of the economy there is potential

for adjustment asymmetries after hit by a shock as well. Therefore, this study has estimated some of these sector specific parameters, which determine the magnitude, and duration of the sectoral response to a shock. The estimates of the underlying parameters clearly indicate that these parameters diverge between the sectors studied and relative price movements have an important role in existing inflation dynamics. Moreover, those sectors each have a specific relation to the rest of the economy in other respects also as the magnitude and duration of the shocks do diverge between the sectors studied. Because the identified parameters rise from the preferences of the agents, the shocks studied could be seen as structural ones as they are related to behaviour of the agents as well. These shocks are hence interpretable and consistent with microeconomic evidence as the previous study of Koskinen and Vilmunen [5] demonstrates. This, in turn, enables us to utilise this DSGE model for a policy analysis that recognize the sector specific features.

Then, concerning the policy implications of this model, we find out that the natural / potential level of output and hence output gaps varies between the sectors studied. As any stabilization as well as structural policy is conditional to the output gap, this fact should be taken into account when practising those policies. These divergences arise from the different preference parameters of the agents. The same applies when dampening the business cycle fluctuations as the magnitude and persistence of adjustment process after hit by a shock do differ between the sectors reflecting the divergence of the underlying parameters.

References

1. Calvo, G. A.: Staggered Prices in a Utility-Maximizing Framework, *Journal of Monetary Economics* 12, 383-398 (1983).
2. Bouakez, H., Cardia, E. and Ruge-Murcia, F.: Sectoral Price Rigidity and Aggregate Dynamics, *European Economic Review*, vol. 65, issue C, 1-22 (2014).
3. Kilponen, J. and Ripatti, A. Labour and product market competition in a small open economy – Simulation results using a DGE model of the Finnish economy. Bank of Finland Discussion Paper 5/2006 (2006).
4. Koskinen, H.: Studies on Money and Labour Market Dynamics and Goods Market Imperfections, Jyväskylä Studies in Business and Economics No. 90. University of Jyväskylä (2010).
5. Koskinen, H and Vilmunen, J.: Implications for Aggregate inflation of Sectoral Asymmetries: Generalizing Woodford, University of Tampere, Tampere Eco nomic Working Papers No 116 (2017).
6. Koskinen, H and Vilmunen, J.: Implications for Aggregate inflation of Sectoral Asymmetries: An Empirical Application, University of Tampere, Tampere Economic Working Papers No 118 (2018).
7. Pfeifer, J.: A Guide to Specifying Observation Equations for the Estimation of DSGE Models, an unpublished manuscript (2015).
8. Tille, C.: The Role of Consumption Substitutability in the International Transmission of Monetary Shocks, *Journal of International Economics* 53, 421-444 (2001).
9. Walsh, C. E.: Monetary Theory and Policy, 3rd edition, Massachusetts Institute of Tecnology. (2010).
10. Woodford, M.: Interest and prices: Foundations of a Theory of Monetary Policy, Princeton, NJ, Princeton University Press (2003).
11. Woodford, M.: Firm-Specific Capital and the New_Keynesian Phillips Curve, International Journal of Central Banking, September 2005, Vol. 1 No. 2 (2005).

Testing for Differences in Forecast-Error Dynamics in Path Forecasts

Andrew B. Martinez*

June 4, 2018

Abstract

Although the trajectory and path of future outcomes play an important role in policy decisions, analyses of forecast accuracy typically focus on individual point forecasts. However, it is important to examine the path forecasts errors since they include the forecast dynamics. We use the link between path forecast evaluation methods and the joint predictive density to propose a test for differences in system path forecast accuracy. We also demonstrate how our test relates to and extends existing joint testing approaches. Simulations highlight both the advantages and disadvantages of path forecast accuracy tests in detecting a broad range of differences in forecast errors. We compare point and path forecasts of inflation across many different model / agency forecasts. Our results show that differences in forecast-error dynamics can play an important role in the assessment of forecast accuracy.

Keywords: GFESM, log determinant, log score, mean square error
JEL classifications: C12, C22, C52, C53

*Department of Economics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, andrew.martinez@economics.ox.ac.uk. This research was conducted in part while the author was a dissertation intern at the Federal Reserve Bank of Cleveland. This research was also supported in part by a grant from the Robertson Foundation (grant 9907422). All numerical results and figures were obtained using OxMetrics 7.2 (OSX/U).

What can drive economic growth in Russia? Mid-term growth scenarios

Svetlana Balashova

RUDN University

sveta.b2@gmail.com

Vladimir Matyushok

RUDN University
vmatyushok@mail.ru

Inna Lazanyuk

RUDN University
liv.rudn@gmail.com

Abstract. This paper examines the prospects of the Russian economy in a medium term. The export-oriented model, driven by high oil prices, providing the growth for the Russian economy before the global crisis, has exhausted itself.
We empirically show that postponement of pension reform and underestimation of the human capital impact on economic growth would lead to a weak economic growth with the average pace no more than 1.8 percent per annum.
We estimate that the target share of investment in GDP can be achieved by a gradual reduction in the real interest rate, increase in industrial production growth rate and technological innovation expenditures.
We show that in medium term main drivers of sustainable economic growth with the pace from 3 to 5 percent per annum are human capital, quality of political institutions, industrial modernization and a significant increase in R&D sector.
JEL: C53, E17, F43
Keywords: Russia, economic policy**,** scenarios, total factor productivity, human capital, drivers of growth, macroeconometric forecasting

1 Introduction

Russia has seen an economic recession during the last five years while the majority of the countries has recovered from the global financial crisis. The export-oriented model, driven by high oil prices, providing the growth for the Russian economy before the global crisis, has exhausted itself. The growth deceleration of the Russian

adfa, p. 1, 2011.

economy has occurred long before the decline in oil prices and the imposition of sanctions in 2014, which is fully understood by the Russian government and academics.

Following the global financial and economic crisis of 2008 - 2009, Russia experienced a period of post-crisis economic recovery, which ended by late 2011. The economic development of Russia in 2011 - 2016 was generally characterized by a gradual weakening in the dynamics of development with an average annual growth rate of 1.2 % against 4.8 % for the period from 2001 to 2010 (see Table 1).

Until 2014, the GDP dynamics was supported by both domestic and external demand. Since 2014 there has been a sharp decline in the fixed assets investment, a decrease in real income, a decline in exports and other negative trends (see Table 1). External factors such as a decrease in energy prices in the world market, the geopolitical tension, and economic sanctions against Russia have led to increased uncertainty and a rapid deterioration in business confidence, negatively affected investment demand and consumer sentiment, caused capital outflow and inflation increase.

Table 1. Key indicators of Russia's socio-economic development in 2011-2017

Indicator	2011	2012	2013	2014	2015	2016	2017
GDP, %	104.3	103.5	101.3	100.7	96.3	99.8	101.4
Consumer price index,% by December	106.1	106.6	106.5	111.4	112.9	105.4	102.5
Industrial production index, %	105.0	103.4	100.4	101.7	96.6	101.3	101.0
Agricultural production index,%	123.0	95.2	105.8	103.5	102.6	104.8	102.4
Investments in fixed assets,%	110.8	106.8	100.8	98.5	89.9	99.8	104.4
Real disposable income,%	100.5	104.6	104.0	99.3	96.8	94.2	98.3
Export of goods, bln. USD	515.4	527.4	523.3	497.8	340.3	281.8	353.0
Import of goods, bln. USD	318.6	335.8	341.3	308.0	194.1	191.6	238.0

Source: Federal State Statistics Service of RF

In 2017 we can see “green shoots” of economic recovery from the last crisis. Efforts on developing infrastructure, the focus on building a digital economy, an increase in industrial and agricultural production, an increase in non-oil export give a base for economic growth expectations. There are several strategies for economic development, in particular, the strategy by the Ministry for Economic Development, the Center for Strategic Research, the Stolypin Club, the RAS Institute for National Economic Forecasts, etc. (see review in [1]). The forecasts are based on different approaches, theoretical models and explicit or implicit assumptions that often remain undisclosed.

In May 2018 newly elected President Putin put forward very ambitious tasks in the socio-economic development of Russia. The Russian government designed a pro-

gram to achieve these goals. The aim of our quantitative analysis is to estimate the expected economic growth under various assumptions. In the target scenario, we rely upon the government objectives regarding main economic drivers.

The contribution of technological progress and human capital accumulation to the economic growth is a classic of economic growth models formulated by Solow and Swan [2, 3]. We rely upon this model and design a set of equations to examine the main building blocks of economic growth in a medium term.

We propose a forecast for the Russian economy, assuming the reasonable time paths of the main drivers. Three scenario forecasts are calculated: basic, assuming the retention of the current model of economic development with moderately favorable external conditions and overcoming the investment decline; second scenario, assuming the retention of the negative demographic trends and the lack of human capital growth; and the target one, under which it is possible to achieve the necessary indicators and to subsequently reach a stable self-sustaining growth path. Econometric tools are used to predict growth drivers.

The main finding of the analysis is that for Russia to sustain a sufficient real GDP growth, it is essential to focus on reforms that drive TFP growth. Modernization of industry, focus on innovations, and the R&D sector enhance are among essential measures. The demographic situation is one of the challenges for future development. It is shown that decline in effective labor led to a decline in expected growth. Acceleration in total factor productivity growth and labor productivity in particular can overcome this risk.

2 Methodological thoughts

2.1 Theoretical model

The neoclassical Cobb-Douglas production function determines the output depending on the capital, effective labor and total factor productivity:

$$Y_t = A_t K_t^{1-\beta} (HC_t AVH_t L_t)^{\beta} \qquad (1)$$

where Y_t is the output, A_t is the total factor productivity, K_t is the accumulated physical capital, HC_t is the human capital per worker, AVH_t is the average annual hours worked by persons engaged. L_t is the total number of workers, presented in the economy. β stands for the labor share of income. Here, the effective labor is the product of human capital per worker HC_t and the total number of hours worked in a certain year $AVH_t \cdot L_t$. Human capital is estimated by the HC index drawn from the PW9.0 database. The assessment methodology is based on the Lee-Barro [4] approach.

The accumulation of physical capital is realized via investment I_t, the previously accumulated capital is depreciated with the depreciation rate δ

$$K_t = (1-\delta)K_{t-1} + I_{t-1} \qquad (2)$$

Equation (1) with help of equation (2) can be rewritten in terms of various growth rates as follows:

$$g_{Y,t+1} \approx g_{A,t+1} + \beta(g_{L,t+1} + g_{AVH,t+1} + g_{HC,t+1}) + (1-\beta)\left[\frac{\frac{I_t}{Y_t}}{\frac{K_t}{Y_t}} - \delta\right] \qquad (3)$$

Here $g_{X,t+1}$ is the growth rate of factor X from t to $t + 1$.

Equation (3) shows what drives the economic growth in this model. Output growth can be achieved through accumulation of capital resources, accumulation of labor resources and productivity growth. Before constructing the growth scenarios, it is necessary to design an empirical model to analyze the dynamics of the main growth drivers.

2.2 Empirical model

Assuming that the relationship (3) is held in long-term and the parameters labor share of income β and depreciation rate δ are constants we examine an empirical model

$$g_{Y,t} = \gamma g_{A,t} + \beta(g_{L,t} + g_{AVH,t} + g_{HC,t}) + \alpha xI_{t-1} / xK_{t-1} + \delta' + \varepsilon_t \qquad (4)$$

with two restrictions $\gamma = 1$, $\alpha = 1 - \beta$. Here $xI = I / Y$, $xK = K / Y$, ε_t is an error term, $\delta = -\delta'/(1-\beta)$.

The total factor productivity (TFP) has the largest direct effect on economic growth. Thus, our primary focus is to design an empirical model to forecast the TFP dynamics. On one side, the TFP is the result of the new knowledge and technologies production, on another side, it depends on the quality of political and economic institutions [5]. TFP also reflects the fluctuations in the market conditions for the goods and services that are produced by the economy. So, assume that there is a "long-run" equilibrium relationship between TFP and considered variables:

$$A_t = C(11)*RD_{t-5} + C(12)*Insitutes_t + C(13)*Export_t + C(14) + \varepsilon_{1t} \qquad (5)$$

where A is the total factor productivity index, RD is the expenditures on R&D, $Export$ stands for the volume of Russian export (the variables are expressed in constant 2010US$ and in logs). $Insitutes$ is a synthetic variable reflecting the quality of public institutions.

R&D expenditures are considered as a proxy of new knowledge and technologies production. We could partly agree with the authors [6], pointing out that the impact of the R&D expenditures should affect the total factor productivity with a lag of

10-15 years. However, due to the short time series for the empirical evaluation of the TFP and considering that modern technologies are being updated much more quickly, we used the lag of five years in equation (4).

To characterize the quality of public institutions, we calculate the arithmetic mean of six indicators of political risks measured by The PRS Group (for more details on variables and the methodology, see the International Country Risk Guide [7]).

There are hundreds of studies that explain the role of export in explaining growth of GDP. Export has a profound impact on economic growth in export-oriented economies [8]. The Russian economy is on the way from the export-oriented commodity economy toward the endogenous sustainable growth model [9]. The share of export in GDP has declined from 44% in 2000 to 26% in 2017. The share of raw materials export, mainly hydrocarbons in total exports is still rather high (around 50% in 2017), but has a tendency to decline since 2011, not only because of the decline in oil prices. Hence, the share of non-raw goods and services in total export has an upward trend. Thus, we assume that raw materials export (denote as $Export\_oil$) is expected to follow the time path of oil prices (denote as Oil), but partly corrected in the next period

$$g_{Export\_oil,t} = C(21)*g_{Oil,t} + C(22)*g_{Export\_oil,t-1} + C(23) + \varepsilon_{2t} \quad (6)$$

We assume that export of non-raw goods and service will increase along with industrial production increase plus 1% expected increase in world demand.

Thus to forecast the TFP dynamics we need to estimate parameters of equations (4-5) and make reasonable assumptions regarding exogenous variables.

The amount of investment in fixed assets and the share of investment in the GDP are endogenous variables of the model.

The growth rate of investment depends on the one hand, on the demand from industry and on the other hand, on the investment climate and the availability of credit resources. Under Western sanctions, foreign money markets are closed to domestic companies and the availability of credit resources in the domestic market plays a major role. It worth to note, that investment in fixed assets is positively and significantly correlated with firms' expenditure on technological innovations. We consider the share of these expenditures in GDP as one of the major indicators of modernization and innovative development.

Thus, we assume that the growth rate of investment can be expressed as follows:

$$g_{I,t} = C(31)*g_{PROM,t} + C(32)g_{HTECHEXP,t} + C(33)RIR_t + C(34) + \varepsilon_{3t} \quad (7)$$

Here $PROM$ is index of production of total industry, $HTECHEXP$ is expenditure on technological innovations, RIR is the lending interest rate adjusted for inflation as measured by the GDP deflator.

In equations (5-7) $C(ij)$ are the model parameters, ε_{it} stand for random components.

Thus, our growth model consists of equations (4)-(7). Various econometric techniques have been used to estimate the required parameters.

3 Quantitative analysis

3.1 Estimation results

To estimate the parameters of the equation (4) we use data from the Penn World Tables 9.0. The period for which data is available is 1993-2014. We use the OLS estimator to assess the parameters. The Wald test is used to test the restrictions, the results show that the restrictions are valid.

The point estimation of the effective labor contribution to the output ($\beta = 0.77$) obtained in the study is higher than the estimates provided by the Penn World Table (0.72 for the period up to 2008, then a decrease to 0.66 by 2014). It worth notice that PWT9.0 reports share of labor compensation in GDP at current national prices.

The estimation of the capital depreciation factor ($\delta = 0.05$) appeared to be higher than in the above source. On the other hand, by the FSSS data, between 2000 and 2016, investments in machinery, equipment and vehicles accounted for 30 to 40% of all investments, and these assets have a depreciation factor of more than 12%.

ADF unit root test shows that time series in equation (5) are I(1), so we use Engle-Granger cointegration test to check series are cointegrated. Since we can reject the null of a unit root in the OSL residuals from (5) (null of no-cointegration) then we cannot reject that the examined variables cointegrate. The estimates from OLS in the static equation (equation 5), although consistent, can be substantially biased in small samples, partly due to serial correlation in the residuals. So we use Fully Modified Least Squares to assess parameters C(11)-C(14) of equation (5). The method modifies least squares to account for serial correlation effects and for the endogeneity in the repressors that results from the existence of a cointegrating relationship [10]). We obtain those estimates of the equation (5) parameters [1]: C(11)=0.16[0.04], which means that a 1% increase in R&D expenditures leads to 0.16% increase in TFP index in 5 years controlling other variables; C(12)=0.44[0.14], a rise in quality of institutions index by 0.1 would rise TFP index by 4.4%; C(13)=0.39[0.06], a rise in real national export by 1% would rise TFP index by 0.4%.

To estimate unknown parameters in equation (6) we use OLS with Newey-West standard errors to deal with autocorrelation. The estimated values can be interpreted in terms of the partial adjustment model. C(21)=0.4[0.05] and C(22)=0.5[0.17] show that short-run elasticity of commodity export on oil price is 0.4, however long-run price elasticity is 0.8.

OLS estimator with Newey-West standard errors is used to assess the unknown parameters of equation (7). The results are as follows: C(31)=0.9[0.3], C(32)=0.2[0.04], C(33)=-0.6[0.2]. Thus, a 1 percentage point increase in industrial production index leads to a 0.9 increase in growth rate of investment; a 1 percentage point increase in technological innovation expenditures leads to a 0.2 increase in growth rate of investment; a 1 percentage point increase in real interest rate leads to a 0.6 decrease in growth rate of investment controlling other variables.

[1] hereinafter standard errors in square brackets

Data used to estimate equations (5)-(7) are retrieved from The World Bank database, The Russian Federal State Statistics Service, The Central Bank of Russia, The International Country Risk database.

Obtained results are used to calculate three growth scenarios in order to quantify the growth impact of effective labor, improvements in investment and the TFP growth. It is necessary to propose reasonable time paths regarding the main exogenous variables of the growth model.

3.2 Forecasting prerequisites

Analyzing the data on the behavior of exogenous variables of the model over the last 5 years, based on analysts' forecasts regarding the oil price, taking into account the monetary policy of the Central Bank of the Russian Federation, and the Government's focus on modernization and economic growth, we determined the following guidelines for the behavior of the factors under consideration (see Table 2).

Table 2. Reasonable Time Path of Exogenes Variables

Exogenous variable	*Baseline*	*Effective labor reduction*	*Target*
The TFP determinants			
Real oil price, 2010 $US per barrel	60 in average	56 in average	As in baseline
Institute index	0.45 the last 5-year average	As in baseline	Reach 0.51 (2010-2014 average) by 2024
R&D expenditure, % of GDP	1.1= the last 5-year average	As in baseline	Reach 1.5 by 2024
Investment determinants			
Real interest rate, %	Reduce by 0.5pp each year	As in baseline	Reduce to 2 by 2019
Technological expenditure, % of GDP	1= the last 5-year average	As in baseline	Reach 1.8 by 2024
The industrial production growth rate	2% per annum	As in baseline	5% per annum
Effective labor determinants			
Human capital growth rate	0.4% per annum	0	Same as baseline
Average annual hours worked by persons engaged	Increase to 1985 hours (2014 level) by 2024	Remain at current level	Same as baseline
Number of persons engaged (in millions)	Remain at the current level due to unemployment reduction and	Decrease to 71.3 by 2024	Increase to 73.3 by 2024 due to the pension reform and migra-

	migration		tion
Population (in millions)	Increase to 148 by 2024	Decrease to 145 by 2024	Increase to 150 by 2024

The constructed scenarios are as follows:

- *Baseline scenario*: a slight increase in all basic indicators in accordance with the prevailing dynamics of recent years and results obtained from equations (5-7);
- *Effective labor reduction*: reduction in working age population according to with FSSS projection, zero growth rate of human capital due to decrease of return on education [11];
- *Target scenario*: a significant increase in investment due to the growth of industrial production and lower real interest rates; an increase in TFP due to improvement in quality of institutions, a gradual increase in R&D expenditure share in GDP and an increase in non-mineral export of ; an increase in effective labor due to the pension reform and a significant increase in public expenditures on human capital.

3.3 Scenario calculations main findings

In all the scenarios, investments in fixed assets have increased as to the previous period, beginning from 2019, yet, at different rates (see Fig. 1). The growth of investments can be achieved by improving the investment climate, expanding the public-private partnerships, reducing the cost of credit resources. However, the growth of investment per se does not guarantee the GDP growth. The need for a more effective use of investments, in particular, through the application of project finance, is essential [12].

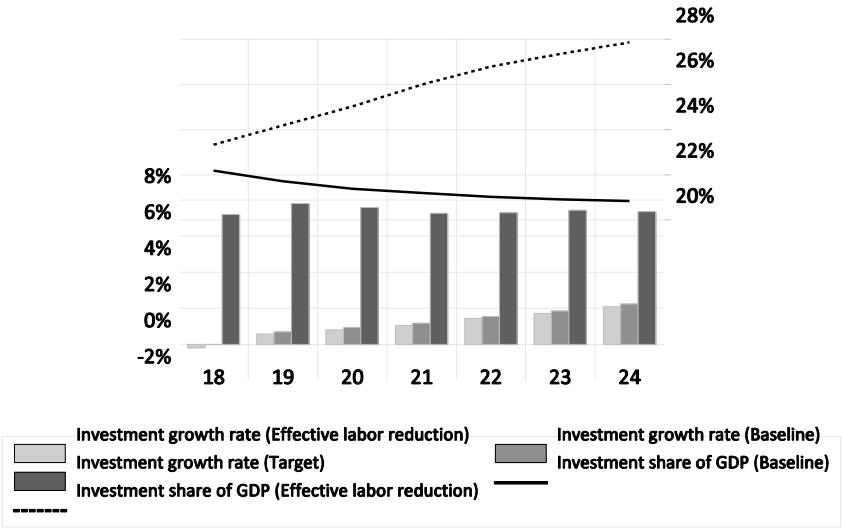


Fig. 1. Dynamics of investments in fixed assets, % as to the previous year (left axis); investment share of GDP (Target and baseline scenarios, right axis). Source: authors' calculation.

The proposed assumptions and estimates allow us to obtain forecast values for the GDP and economic growth in the medium term. Econometric models estimated on relatively short time series do not allow to make long-term forecasts. Furthermore, the impending technological revolution is likely to drastically change the conditions for the economic systems functioning [13].

The results of scenario calculations and evaluation of key macroeconomic indicators according to scenarios are presented in Fig. 2 and Table 3.

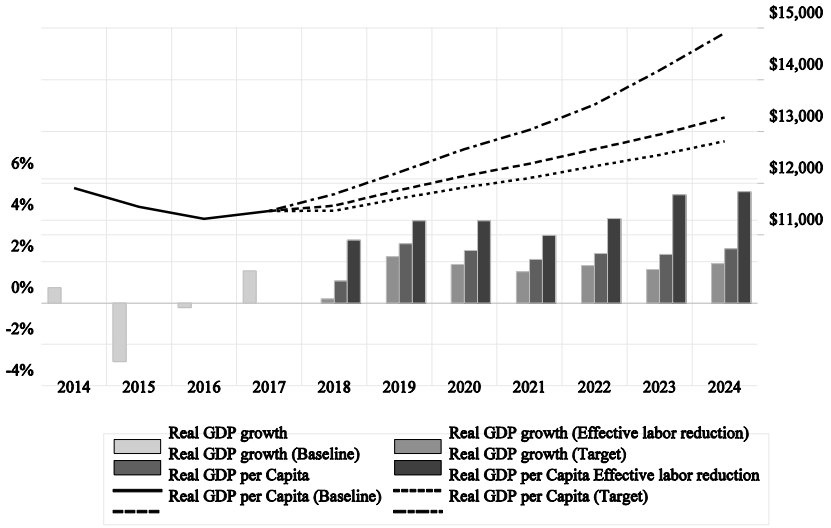


Fig. 2. GDP per Capita (in constant 2010 US$, right axis) and real GDP growth rates (%, left axis) by scenario options. Source: WDI, authors' calculation

The baseline scenario estimates are close to those by the World Bank [14] and the IMF [15]. This scenario is based, on the one hand, on preserving the current trends in the economy. On the other hand, it requires the growth of industrial production, the growth of non-primary exports, the growth of human capital, and the preservation of the number of people employed in the national economy. We called the scenario baseline not because it does not require any effort to be implemented. Signs of Dutch disease are still inherent in the Russian economy [16], and it does not allow us to count on high rates of economic growth even under moderately favorable external conditions.

Table 3. Scenario-based estimates of the key macro-economic indicators

Indicator	**2018-2020**	**2021-2022**	**2023-2024**
Real GDP growth, %			
Effective labor reduction	1.4	1.7	1.8
Baseline	2.2	2.2	2.5
Target	3.6	3.7	5.3
Gross capital formation, % of GDP			
Effective labor reduction	21	21	21
Baseline	22	21	21
Target	24	26	27
GDP per capita, thousand 2010 US dollars			
Effective labor reduction	11.700	12.200	12.700
Baseline	11.800	12.500	13.100
Target	12.200	13.300	14.500

The second scenario shows relative importance of labor recourses in economic growth. Ageing of the population, low returns from the level of education among young people, low mobility of the labor force in modern Russia can lead to a reduction in the quantity and quality of the workforce. Despite the growth in investment, the same as in the baseline scenario (see Fig. 1), lower economic growth rates are expected (see Fig. 2) and a noticeable lag in GDP per capita from the baseline scenario towards the end of the forecast period (see Fig. 2 and Table 3).

The target scenario can be achieved due to the combination of various growth drivers. The TFP growth (up to 4% per annum) can be accomplished due to the higher R&D expenditures growth as against the GDP growth, as well as due to the growth in non-resource exports while maintaining a moderately favorable environment of the world raw materials market. In boosting the TFP growth the role of public institutions is essential. The projection of growth of institutions quality index is based on the guess in improving the quality of public service delivery, an efficiency of same by dismantles bureaucratic obstacles to business and creates an environment-friendly to investment and business, improving the legislation towards an increase in the rule of law index, and a greater level of business confidence. The contribution of public institutions in an increase of R&D sector is vital too. As it was shown [17], direct and indirect public expenditures on R&D stimulate business to enlarge R&D expenditures. Also, the target scenario implies a meaningful increase in investments due to the gradual reduction of the real interest rate, a considerable increase in demand from industry, the investment climate changes. As a result, the share of investments in fixed assets in the GDP is expected to increase up to 27% by 2024. The target scenario realization is crucial to achieving a per capita GDP growth rate of more than 3% a year, which is necessary to solve the problem of the population's welfare gradual improvement.

4 Conclusion

The aim of this study is to analyze the factors affecting the economic growth in Russia. As the basic model, the Solow-Swan model is adopted, which assumes that output growth is provided in three ways: capital accumulation, effective labor and productivity growth.

To simulate the dynamics of total factor productivity and capital accumulation, a system of econometric equations is proposed. It includes traditional growth drivers, and specific factors we believe to be the drivers of the expected growth, such as the quality of political and economic institutions and R & D expenses.

Our findings suggest that human capital along with the creation of conditions for keeping the number of employed in the economy can provide a meaningful contribution to economic growth. The postponement of pension reform, neglecting human capital investment, low mobility of labor force in Russia and aging of the population would lead to a weak economic growth.

We detected that there is a need for a significant increase in the total factor productivity (up to 4% per annum) on a new technological basis, due to a greater efficiency of R&D expenditures and the effectiveness of the national innovation system, increasing investment in innovative high-tech production, diversifying exports, better functioning of political and economic institutions. This will significantly increase the economic health and, on the basis thereof, - the incomes of those employed in the economy, and later - in the social and budgetary spheres. This, in turn, will stimulate consumer demand, the growth of which will facilitate warming up the economy and, with good managerial and project solutions available, achieving the required rates of economic growth.

5 Acknowledgements

Research reported in this paper was supported by Russian Humanitarian Foundation Grant No. 16-02-00375.

6 References

1. Matyushok, V.: On strategies of economic growth in Russia. Rudn J. Econ. 4, 7–19 (2016)
2. Solow, R.: A Contribution to the Theory of Economic Growth. Q. J. Econ. 70, 65–94 (1956)
3. Swan, T.: Economic Growth and Capital Accumulation. Econ. Rec. 32, 334–361 (1956)
4. Barro, R.J., Lee, J.W.: A new data set of educational attainment in the world, 1995-2010. J. Dev. Econ. 104, 184–198 (2013)
5. Acemoglu, D., Johnson, S., Robinson, J.A.: Institutions as a fundamental cause of long-term growth. In: Handbook of Economic Growth. p. 1138 (2005)

6. Apokin, A.Y., Ipatova, I.B.: Components of total factor productivity of the Russian economy with respect to other countries of the world: The role of technical efficiency. Stud. Russ. Econ. Dev. 28, 15–21 (2017). doi:10.1134/S1075700717010026
7. International Country Risk Guide, https://www.prsgroup.com/explore-our-products/international-country-risk-guide/
8. Kristjanpoller, W., Olson, J.E., Salazar, R.I.: Does the commodities boom support the export led growth hypothesis? Evidence from Latin American countries. Lat. Am. Econ. Rev. 25, 1–13 (2016). doi:10.1007/s40503-016-0036-z
9. Shirov, A.A.: Structural aspects of long-term economic dynamics. In: Economic policy: choice of strategy and issues of tactics (2017)
10. Phillips, P.C.B.: Fully Modified Least Squares and Vector Autoregression. Econometrica. 63, 1023–1078 (1995). doi:10.2307/2171721
11. Arkhipova, M., Egorov, A., Syrotin, V.: Return to scooling in Russia and Ukraine: Comparitive analysis. Appl. Econom. 47, 100–122 (2017)
12. Abramova, E., Apokin, A., Belousov D., Mikhailenko, K., Penukhina, E., Frolov, A.: Future of Russia: Macroeconomic scenarios in the global context. Foresight Russ. (2013)
13. Schwab, K.: The Fourth Industrial Revolution. Foreign Aff. 12, (2015)
14. World Bank Group: Russia Economic Report. (2017)
15. IMF: Seeking Sustainable Growth: Short-Term Recovery, Long-Term Challenges. (2017)
16. Mironov, V. V, Petronevich, A. V: Discovering the signs of Dutch disease in Russia. Resour. Policy. 46, 97–112 (2015)
17. Balashova, S.: The impact of public R&D policy on business-funded R&D (case of OECD countries). Appl. Econom. 38 (2), 64–82 (2015)

Determining the Cointegration Rank Using a Residual-based Procedure.

By Antonio Aznar. University of Zaragoza

Abstract

In this paper, we derive a new test of cointegration based on a residual based approach. In each step we determine either the cointegration rank or the set of common trends to be used in the following step. We derive the probabilities of choosing the correct rank and of overidentifying and underidentifying that rank. We present results from a simulation experiment to asses the performance of the test with finite samples. Then, we apply the test to quarterly data of four variables of the Danish economy.

JEL Classification: C32

Keywords: Cointegration rank, sequential approach, Augmented Dickey-Fuller test, simulation experiment, empirical application.

1. Introduction

In this paper, we propose a procedure to test the cointegration rank of a multivariate dynamic system. In their seminal paper, Engel and Granger (1987) used the Dickey-Fuller test to determine the existence of cointegration. But they restrict their approach to only one cointegration relation and since then there has been only a few attempts to extend that procedure to a more general setting with n variables and r cointegration relations. For an example see the GBH test in Gomez-Biscarri and Hualde (2015).

The most well-known procedure for the general setting has been that proposed by Johansen (1988, 1991 and 1995) based on maximum-likelihood inference on vector autoregressive error correction (CVAR) models. The Johansen's test is a sequential procedure such that in each step of the process it considers a null of less cointegration against an alternative of more cointegration. A likelihood ratio statistic is calculated using the characteristic roots of a given matrix defined in terms of the second moments of the variables of the so-called Reduced Rank Regression. The difference between two consecutive likelihood ratio statistics is that when defining the second one the largest root used in the previous one is excluded.

The cointegration test we propose in this paper is also a sequential procedure. But it is a residual based test such that in the different steps of the process we calculate the values taken by the Augmented Dickey-Fuller (ADF) statistic for a given set of regressions and the decision is based on the maximum of these values. In each step, we also consider a null hypothesis of less cointegration against an alternative hypothesis of more cointegration. However, there are two differences with respect the Johansen's approach. In our approach, the null hypothesis is composed while the alternative is simple. In the Johansen's approach, the null hypothesis is simple and the alternative is composed. Secondly, there is a difference in the location of the null hypothesis; in the Johansen's approach, the null hypothesis of a zero cointegration appears in the first step of the process; in our approach, the null hypothesis of a cero cointegration rank corresponds to the last step. It is worth remarking that, in each step of the process, we determine either the cointegration rank- if the null hypothesis is rejected- or the set of common trends to

be used in the following step-if the null hypothesis is not rejected.

The paper is organized as follows. In Section 2 we present The model and some preliminary results. Section 3 is dedicated to describe the new test. The simulation results are presented in Section 4. In section 5 we present the results of an empirical application and the paper ends in Section 6 with the main conclusions**.**

2. Model and some Preliminary Results

We consider an n-dimensional I(1) system y_t that can be written as

$$\Delta y_t = \delta + u_t \qquad (1)$$

Where δ is an n-dimensional vector of constants representing the mean of the stationary process; u_t is a linear zero-mean n-dimensional vector I(0) process that can be written as

$$u_t = \Psi(L)\varepsilon_t \qquad (2)$$

Where $\Psi(L) = I_n + \Psi_1 L + \Psi_2 L^2 + \ldots$. ε_t is i.i.d. with

$$E\varepsilon_t = 0, E\left(\varepsilon_t \varepsilon_t^{'}\right) = \Omega, \Omega \text{ positive definite}$$

We assume that $\Psi(L)$ satisfies two conditions

i. $\{\Psi_j\}$ is one-summable
ii. $\Psi(1) \neq 0$

As it can be seen in Hayashi (2000), the first condition makes u_t being a stationary and ergodic process.

The long-run covariance of the I(0) system is

$$\psi(1)\Omega\Psi(1)' \qquad (3)$$

In levels, y_t can be written as

$$y_t = y_0 + \delta t + u_1 + u_2 + ...u_t \qquad (4)$$

Then, consider the following result associated with the Beveridge-Nelson decomposition

$$\Psi(L) = \Psi(1) + (1-L)\alpha(L),$$
$$\alpha(L) = \sum_{j=0}^{\infty} \alpha_j L^j, \alpha_j = -\left(\Psi_{j+1} + \Psi_{j+2} + ...\right)$$

Using this result we can write (2) as

$$u_t = \Psi(1)\varepsilon_t + \eta_t - \eta_{t-1}, \eta_t \equiv \alpha(L)\varepsilon_t \qquad (5)$$

Where $\eta(L)$ is a well-defined covariance-stationary process. Substituting (5) into (4) we obtain the multivariate version of the Beveridge-Nelson decomposition

$$y_t = \delta t + \Psi(1)\left(\varepsilon_1 + ... + \varepsilon_t\right) + \eta_t + \left(y_0 - \eta_0\right) \qquad (6)$$

We assume that there exists a $p \times r$ matrix, β, such that

$$\beta'\Psi(1) = 0 \quad and \quad \beta'\delta \neq 0$$

β is the matrix of cointegration vectors and r is the cointegration rank.

The corresponding CVAR model is

$$\Delta y_t = \mu_0 + \alpha\left(\rho_1 t + \beta' y_{t-1}\right) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \qquad (7)$$

Where ρ_1 is a vector of r constants.

For a particular order of the n variables, let $\hat{\tau}_{k-1,2,..l}$ the value of the Augmented Dickey-Fuller(ADF) statistic calculated with the residuals of the regression of y_{kt} on $y_{1t}, y_{2t}, \ldots y_{lt}, t$ with k>l and $l \leq n-1$. Note that when l=1 the number of regressions is equal to the $\binom{n}{2}$ non-redundant pairs of variables while if l>1 the number of regressions will be n-l. Let $\hat{\tau}^*_{k-1,2,..l}$ be defined as

$$\hat{\tau}^*_{k-1,2,..l} = \max_k \hat{\tau}_{k-1,2,..l} \qquad (8)$$

Note that, if one of these regressions is a cointegrating regression, then the corresponding ADF statistic satisfies

$$\lim_{T\to\infty} \Pr\left(\hat{\tau}_{k-1,2,..l} < -K\right) = 1$$

For any arbitrarily large positive K. On the other hand, if one of these regressions is not a cointegration relation then there exists a real number, $v_{l+1}(\alpha)$, for which the corresponding ADF statistic satisfies

$$\lim_{T\to\infty} \Pr\left(\hat{\tau}_{k-1,2,..l} < v_{l+1}(\alpha)\right) = \alpha$$

As a consequence of these results we can conclude saying that if, for a given l, there are n-l regressions that are cointegrated, then

$$\lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-1,2...l} < -K\right) = 1 \qquad (9)$$

If, for any given l, there exists, at least, one regression that is not a cointegration relation, then

$$\lim_{T\to\infty} \Pr\left(\hat{\tau}^{*}_{k-1,2\ldots l} < v_{l+1}(\alpha)\right) = \alpha \qquad (10)$$

3. Cointegration Test

It is a sequential test with (n-1) steps. In the first step, we test for the presence of, at least, two common trends against the alternative that there exists only one common trend. In the second step, we test for the presence of, at least, three common trends against the alternative hypothesis of two common trends .The process follows until we get up the final step in which, under the null hypothesis, the number of common trends is n and under the alternative hypothesis the number is n-1.

Step 1. In this step, the null hypothesis we test is $H_0 : r \le 2$ while the alternative specifies $H_A : r = r - 1$. Consider the regressions corresponding to the $\binom{n}{2}$ non-redundant pairs of variables

$$y_{kt} = \alpha + \delta t + \phi y_{lt} + u_t \quad \text{k>l, l=1,2,..n-1}$$

Then, we calculate the corresponding $\binom{n}{2}$ ADF values and the null hypothesis is rejected when

$$\hat{\tau}^{*}_{k-l} \le cv_2(\alpha)$$

If the null hypothesis is rejected then the process stops and we conclude that the cointegration rank is n-1. If the null hypothesis is not rejected we go to the next step of the process assuming that the two first common trends are the two variables of the regression to which corresponds the largest ADF value.

Step2. In this step, we test the null hypothesis $H_0 : r \leq n-3$ against the alternative $H_A : r = n-2$. Without loss of generality, suppose that the largest ADF value in Step 1 corresponds to the regression with $y_{kt} = y_{2t} \quad and \quad y_{lt} = y_{1t}$. Then, consider the following n-2 regressions

$$y_{kt} = \alpha + \delta t + \phi_1 y_{1t} + \phi_2 y_{2t} + u_t \quad \text{k=3,4...n.}$$

We calculate the corresponding n-2 ADF values and the null hypothesis is rejected when

$$\hat{\tau}^*_{k-1,2} \leq cv_3(\alpha)$$

If the null hypothesis is rejected then, the testing process stops and we conclude that the cointegration rank is n-2. If the null hypothesis is not rejected we proceed to the following step …until we arrive to the final step.

Step (n-1). In this case, the null hypothesis tested is that the cointegration rank is 0, r=0, against the alternative that it is one, r=1. We estimate only the following regression

$$y_{nt} = \alpha + \delta t + \phi_1 y_{1t} + \phi_2 y_{2t} ..... + \phi_{n-1} y_{n-1t} + u_t$$

If the null hypothesis is rejected, then the conclusion is that the cointegration rank is one. If it is not rejected, then the conclusion is that there is no cointegration.

Let $\hat{r}$ be the cointegration rank estimated applying the sequential process we have just described. The properties of this estimate are given in the following theorem

Theorem 1. Let $r \in \{0,1,2,....n-1\}$ be the true cointegration rank. Then

i. $$\lim_{T\to\infty} \Pr\left(\hat{r}=r\right)=\left(1-\alpha\right)^{n-1-r} \quad (11)$$

ii. $$\lim_{T\to\infty} \Pr\left(\hat{r}=k\right)=\left(1-\alpha\right)^{i-1}\alpha$$, k=r+i, i=1,2,..n-r-1, r<k (12)

iii. $$\lim_{T\to\infty} \Pr\left(\hat{r}=k\right)=0$$, k=0,1,2….r-1, r>0 (13)

Proof: First, we show (11) for r=n-1. We have seen that $\hat{r}=n-1$ when, in the first step, the null hypothesis $H_0: r \le n-2$ is rejected. That is, when

$$\hat{\tau}^*_{k-l} \le cv_2\left(\alpha\right)$$

The result follows because if r=n-1 using (10) we have

$$\lim_{T\to\infty} \Pr\left(\hat{r}=n-1\right)=\lim \Pr\left(\hat{\tau}^*_{k-l} \le cv_2\left(\alpha\right)\right)=1$$

Next, we show (11) for r=n-2. The estimated rank is equal to n-2 when the null hypothesis $H_0: r \le n-2$ is not rejected in the first step, and the null hypothesis $H_0: r \le n-3$ is rejected in the second step. That is, $\hat{r}=n-2$ when

$$\hat{\tau}^*_{k-l} > cv_2\left(\alpha\right) \; and \; \hat{\tau}^*_{k-1,2} < cv_3\left(\alpha\right)$$

The result follows because if r=n-2, (9) implies that

$$lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-l} > cv_2\left(\alpha\right)\right)=\left(1-\alpha\right)$$

On the other hand, using (10) we have

$$lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-1,2} \le cv_3\left(\alpha\right)\right)=1$$

The result for n-j, j=3,4,..n can be derived using similar arguments.

Next, we show (12). Let r=n-j. The estimate $\hat{r} = n - i$ is obtained when

$$\hat{\tau}^*_{k-l} > cv_2(\alpha), \hat{\tau}^*_{k-1,2} > cv_3(\alpha), ...\hat{\tau}^*_{k-1,2..i-1} > cv_i(\alpha)$$
$$and\ \hat{\tau}^*_{k-1,2..,i} \leq cv_{i+1}(\alpha)$$

The result follows because, for i<j, (9) implies that

$$\lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-l} > cv_2(\alpha)\right) = (1-\alpha)$$
$$\lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-1,2..l} > cv_{l+1}(\alpha)\right) = (1-\alpha), l = 2,..,i-1$$

And

$$\lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-1,2,..,i} \leq cv_{i+1}(\alpha)\right) = \alpha$$

Next, we show (13) for k=r-j, j=1,2,..r, r>0. Let r=n-i and j>i. We obtain $\hat{r} = r - j$ when the following results jointly hold

$$\hat{\tau}^*_{k-l} > cv_2(\alpha), \hat{\tau}^*_{k-1,2} > cv_3(\alpha), ...\hat{\tau}^*_{k-1,2..n-r+j-1} > cv_{n-r+j}(\alpha)$$
$$and\ \hat{\tau}^*_{k-1,2..n-r+j} \leq cv_{n-r+j+1}(\alpha)$$

The result follows because using (9) and (10) we have

For l<i: $\lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-1,2,..,l} > cv_{l+1}(\alpha)\right) = 1-\alpha$

For l≥i: $\lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-1,2,..,l} > cv_{l+1}(\alpha)\right) = 0$

$$\lim_{T\to\infty} \Pr\left(\hat{\tau}^*_{k-1,2,..,l} \leq cv_{l+1}(\alpha)\right) = 1$$

It is easy to show that, for a given cointegration rank, r=n-j, the sum of probabilities of choosing the correct cointegration rank and of over identifying that rank is equal to one. Just write

$$\Pr\left(\hat{r}=n-j\right)+\Pr\left(\hat{r}=n-j+1\right)+..+\Pr\left(\hat{r}=n-1\right)=$$
$$\left(1-\alpha\right)^{j-1}+\alpha+\left(1-\alpha\right)\alpha+.....+\left(1-\alpha\right)^{j-2}\alpha$$

And the result follows because

$$\left(1-\alpha\right)^{j-1}+\left(1-\alpha\right)^{j-2}\alpha=\left(1-\alpha\right)^{j-2}$$

4. Simulation Results

In this section we present the results of a Monte-Carlo simulation study to assess the performance of the proposed test in an scenario with small samples.

We assume a three variable model. First, we generate three autoregressive processes as follows

$$v_{jt}=\rho_j v_{jt-1}+\varepsilon_{jt} \quad j=1,2,3 \qquad (11)$$

Where $\left|\rho_j\right|<1$ and the $\varepsilon's$ are white noise independent among them. Then, we generate the values of the three variables. When there is no cointegration the process we follow to generate the data is

$$\Delta y_{jt}=\delta_j+v_{jt} \quad j=1,2,3 \qquad (12)$$

When the cointegration rank is one, the first two variables are generated by (12), while the third variable is generated as

$$y_{3t}=\delta_{13}+y_{1t}+y_{2t}+v_{3t} \qquad (13)$$

When there are two cointegration relations, the two first variables are generated by (12) while the last one is generated by

$$y_{2t} = \delta_{22} + y_{1t} + v_{2t}$$
$$y_{3t} = \delta_{23} + y_{1t} + v_{3t} \qquad (14)$$

This is the model we have introduced in Section 2. As we indicated, in this model the variables have a time trend and these trends are not cointegrated. We consider the following set of values: $\delta_j = .2 \; and \; \delta_{13} = \delta_{22} = \delta_{23} = .5$. When the errors are white noise we set $\rho_j = 0$ while, if we assume autoregressive errors, the value is $\rho_j = .8$. Additionally, we consider two other generating models, which are restricted versions of that model, the first with no deterministic terms, that is, $\delta_j = 0 \; and \; \delta_{13} = \delta_{22} = \delta_{23} = 0$ and the other one in which the variables have time trends that are cointegrated, that is $\delta_j = .2 \; and \; \delta_{13} = \delta_{22} = \delta_{23} = 0$. We do not provide the results of these two models because they coincide with those corresponding to the general model.

The results are presented in Table 1-white noise errors- and in Table 2-autoregressive errors-, for four different sample sizes, T=50, 100 200,500, and different values of r. The number of lags in the ADF test are chosen using the BIC. The cells show the proportion of 5,000 replications where the estimated rank obtained by the new test is $\hat{r}$,given the correct rank r, so that the block in the main diagonal represent the percentage of "correct answers" whereas the rest of blocks indicate proportion of "mistakes". We use the 5% significance level.

Examining the results for the white noise case we see that the performance of our test matches the theoretical properties in Theorem 1. When r=n-1=2 the test correctly estimates the rank

in 100% cases showing evidence of (11). When r<2 that percentage decreases, the worst results corresponding to r=0 as it might be expected from (11). We observe also that the probability of overidentifying is not zero; it is a decreasing function of the distance with respect the true cointegration rank showing evidence of (12). Finally, it can be seen that the probability of under identifying is zero.

Table 1. Performance of the Test. White Noise Error

$\hat{r}$	n/r	2	1	0
	50	.971	.115	.021
2	100	.999	.084	.015
	200	1.00	.060	.009
	500	1.00	.012	.010
	50	.007	.860	.162
1	100	.001	.912	.108
	200	.000	.940	.095
	500	.000	.951	.084
	50	.022	.025	.817
0	100	.000	.004	.877
	200	.000	.000	.896
	500	.000	.000	.906

The innovation vector v_t is generated from $v_{jt} = \varepsilon_{jt}$, j=1,2,3, with Gaussian white noise ε_t such that $E(\varepsilon_t) = 0, Var(\varepsilon_t) = I_3$.

If we pay attention to the case with autoregressive errors, Table 2, we have similar conclusions but now a larger sample is needed. At least 200 observations are required to replicate the results with white noise and 100 observations.

Table 2. Performance of the Test. Autoregressive Error

$\hat{r}$	n/r	2	1	0

	50	.603	.018	.017
2	100	.575	.012	.011
	200	.998	.014	.013
	500	1.00	.002	.012
	50	.241	.211	.157
1	100	.135	.428	.102
	200	.000	.954	.085
	500	.000	.980	.084
	50	.156	.771	.825
0	100	.290	.560	.887
	200	.002	.032	.902
	500	.000	.000	.904

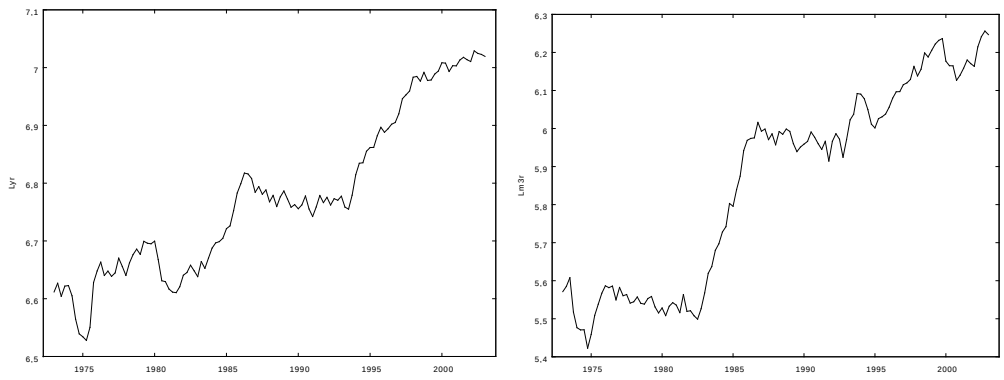
The innovation vector v_t is generated from $v_{jt} = .8v_{jt-1} + \varepsilon_{jt}$ j=1,2,3, with Gaussian white noise ε_t such that $E(\varepsilon_t) = 0, Var(\varepsilon_t) = I_3$.

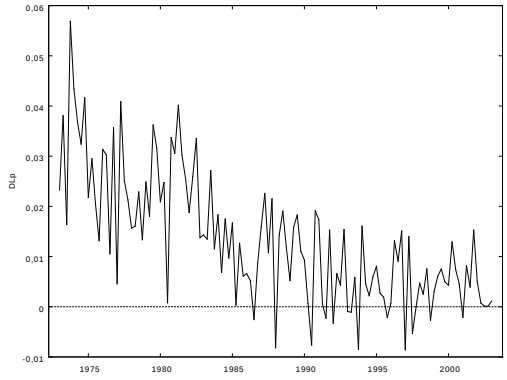
5. Empirical Application

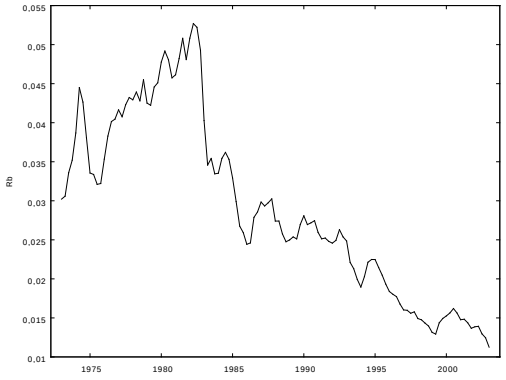
We use the quarterly data of four variables of the Danish economy, from 1973:1 to 2003:1,reported in Juselius (2006).

The data vector is $x_t = (Lyr_t, Lm3r_t, DLp_t, Rb_t)$, where Lyr_t is the log of the real Gross National Product, $Lm3r_t$ is the log of real M3, DLp_t is the rate of inflation and Rb_t is the long-term government bond rate.

Figure 1 Graphs of the variables







Observing the graphs, it is seen that the two first variables, $Lyr_t \ and \ Lm3r_t$, have an increasing temporal pattern while the last two, $DLp_t \ and \ Rb_t$ have a decreasing pattern. Hence, when defining the Dickey-Fuller test we will consider a model with a constant and a time trend and when examining the cointegration relations we will assume that the time trends are not cointegrated, case four of Johansen.

First, let us examine the values of the Dickey-Fuller statistic for the four variables reported in Table 3.

Table 3. Test of the order of integration

Variable	ADF
Lyr	-2.72
Lm3r	-2.53
DLp	-3.72
Rb	-2.42

The values of the ADF clearly indicate to accept that the four variables are I(1).

Next, in Table 4 we present the ADF values of the $\binom{4}{2} = 6$ non-redundant pairs of variables.

Table 4. ADF values in the first step

Model	$\hat{\tau}_{k-l} \quad cv_2\left(\alpha = .05\right) = -3.8$
Lyr-Rb	-2.70
Lm3r-Rb	-1.47
Lm3r-Lyr	-1.87
Dlp-Rb	-3.98
Dlp-Lyr	-6.04
Dlp-Lm3r	-12.52

It is seen that the largest ADF value corresponds to the pair Lm3r-Rb. Since this value (-1.47) is larger than the critical point we decide to maintain the null hypothesis and accept that the two first common trends are Lm3r and Rb.

Next, we examine the ADF values, reported in Table 5, which correspond to the regressions with two common trends in the second step.

Table 5. ADF values in the second step

Model	$\hat{\tau}_{k-1,2} \quad cv_3\left(\alpha = .05\right) = -4.16$
Lyr-Lm3r,Rb	-3.40
Dlp-Lm3r,Rb	-12.76

Since the largest value (-3.40) is larger than the critical point we do not reject the null hypothesis and proceed to the final step of the process.

The ADF value for the last regression with three common trends can be seen in Table 6.

Since this value (-12.46) is smaller than the critical point we conclude that the cointegration rank is one.

Table 6. ADF values of the third step

Model	$\hat{\tau}_{k-1,2,3} \quad cv_4(\alpha = .05) = -4.49$
Dlp-Lm3r,Rb,Lyr	-12.76

Using the Johansen's procedure with two lags that is the number provided by the AIC criterion and using the case of restricted trend the results are as follows

Table 6. Johansen's Results

Rank	Eigenvalues	Test Statistic(p value)
0	0.369	84.18(0.0003)
1	0.118	29.30(0.54)
2	0.077	14.25(0.64)
3	0.038	4.64(0.65)

We obtain the same conclusion that the cointegration rank is one.

6. Conclusions

In this paper, we have derived a new cointegration test. It is a sequential test such that in each step of the process we calculate the values taken by the ADF statistic for the set of regressions corresponding to that step and the decision is based on the maximum of these values. We have shown that in the j-th step of the process the null hypothesis that is tested is that the cointegration rank is equal or smaller than n-j-1, while the alternative hypothesis is that the rank is n.j. If the null hypothesis

is rejected, then we conclude that the cointegration rank is n-j; if the null is not rejected we specify the set of j+1 common trends to be used as regressors in the following step. These common trends are the j+1 variables of the regression to which corresponds the highest ADF value.

In Theorem1 of Section 3, we have obtained the asymptotic probability of choosing the correct cointegration rank, the probability of over identifying that rank and the probability of under identifying it. We have shown that the first probability converges to a value that is lower the lower is the cointegration rank; the probability of over identification converges to a value that depends negatively on the distance between the rank estimated and the true cointegration rank and, finally, the probability of under identifying converges to zero.

In Section 4, we have presented the results of a Monte-Carlo simulation experiment where we have assumed a model with three variables which have linear trends that are not cointegrated. It is shown that the new test has a good performance in the sense that it matches the results of Theorem 1 and because the probability of choosing the true cointegration rank is high. An important result we have obtained is that the results are robust with respect other generating processes like, for example, a model with no deterministic terms or a model in which the variables have time trends that are cointegrated.

The last section of the paper has been dedicated to present the results of an empirical application. The data employed is the quarterly information of the Danish economy used in Juselius(2006). We limit the application to four of the variables: real output, real stock of money, rate of inflation and long-term

bond rate. The graphs of the variables show that, at least, two of them have increasing patterns justifying the inclusion of the time trend in the ADF regressions. We have shown how to implement the different steps of the process and the conclusion is that there exists a cointegration relation.

References

Engel, R. F. and Granger, C. W. J. (1987): Cointegration and error correction model. Representation, estimation and testing. Econometrica, 55, 251-276.

Gomez-Biscarri, J. and J. Hualde (2015): Regression-based analysis of cointegration systems. Journal of Econometrics, 186, 32-50.

Hayashi, F. (2000): Econometrics. Princeton.

Johansen, S. (1988): Statistical analysis of cointegration vectors. Journal of Economics Dynamics and Control, 12, 231-254.

Johansen, S. (1991): Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica, 59, 1551-1580.

Johansen, S. (1995): Likelihood based inference in cointegrated vector autoregressive model. Oxford University Press, Oxford.

Juselius, K. (2006): The Cointegration VAR Model: Methodology and Applications. Oxford University Press.

Phillips, P.C.B. and Ouliaris S. (1990): “Asymptotic properties of residual based testsfor cointegration”, Econometrica, 58, 165-193

Point Function Analysis and a Hypothesis on the Origin of Quantum Mechanics

Bernd Burchard[1][0000−0002−2272−1832]

45276 Essen, Germany bernd.burchard@b-burchard.de

Abstract. A modified calculus capable of handling point-like functions of finite value and zero diameter is proposed. Such point-like functions might follow the Schrödinger-Dirac-Equation of quantum mechanics and the corresponding statistics. The number of Eigen-values corresponds to the number 48 of the fermionic elementary particles known from the standard model of elementary particle physics.

Keywords: Dirac equation · analysis · standard model

1 Introduction

The handling of point-like particles is connected with some severe philosophical problems [1] in the calculus. The first one who questioned the principle "natura non facit saltus" (in English: "nature does not make jumps") was Niels Bohr [2, 3]. Today different methods and approaches are known to define general functions describing singularities. To describe those points, Dirac [4] introduced the Dirac distribution $\delta_D(s) = \{1 \text{ for } s = 0; 0 \text{ for } s \neq 0\}$ for example. Those attempts include new analysis methods with modified derivations. The common problem to be solved of all of theses theories is how to maintain the function smoothness to be able to apply conventional analysis methods. The distribution theory [5, 6] for example uses smooth continuous test functions. A significant problem in distribution theory is multiplication [7]. Therefore distributions do not form an algebra without further measures. A first corresponding exemplary extension of distribution theory is the micro-local analysis [8, 9]. Its basic idea goes back to the properties of the Fourier transform. If f is an integrable compactly supported function, one can tell whether f is smooth by looking at the behavior of its Fourier transform $\hat{f}(\xi)$ (that is smooth, even analytic) when $\xi \to \infty$. Therefore it stays within the smoothness boundary of conventional analysis. A second example is the theory of hyperfunctions [10, 11] that might be seen as a special type of micro-local analysis [15]. Furthermore the wavelet analysis is used in some publications [16, 17] to solve the continuity problem of point like singularities. Other examples of modified analysis methods are the non-standard-analysis (NSA) [12] and the smooth-infinitesimal-analysis (SIA) [13, 14]. All these methods understand the derivation of a function $f(s)$ at a certain location s as a property of the function $f(s)$ at the location s. In contrast to this the derivation of a function $f(s)$ is a property of its environment at the

location s in this text. This eliminates the need for smoothness. The key element is the introduction of a finite resolution $\epsilon \in \mathbb{R}$. The finite resolution ϵ is valid for the observer not for the mathematical system as a whole. With relation to the space of complex numbers $\mathbb{C}$ for example this means that there shall be a minimum radius $\epsilon \in \mathbb{R}$ the observer is unable to distinguish from 0. Therefore the observer can not distinguish any complex number $\alpha \in \mathbb{C}$ with $|\alpha| < \epsilon$ from 0 and judges the equation $\alpha = 0$ to be right. I mark such a judgment with the equal sign $=$. But there is only one true $\alpha \equiv 0$ and an infinite number of other $\alpha \in \mathbb{C}$ inside the area of $|\alpha| < \epsilon$ and $\alpha \not\equiv 0$. In the following, we will investigate the properties of unknown functions. We know the exact function value $f(s)$ only for those locations s_i used as a measurement point s. Each measurement is able to evaluate the function value $f(s)$ at one element s_i of the definition set $\mathbb{V}$ of the unknown function $f(s)$. Therefore, all statements regarding such unknown functions $f(s)$ are only hypotheses. This simple trick leaves the set of complex numbers $\mathbb{C}$ intact and avoids the need to introduce infinitesimals. Our objetive is to avoid the indivisible problem of the standard calculus, as pointed out by Leibnitz [18–20]. In contrast to Diracs delta function [21], we define a finite point function $\delta(s)$ of a parameter $s \in \mathbb{C}$ to avoid this indivisible problem of the standard calculus [18–20]. Two mathematical objects shall be equal ($=$) if all observable properties do not differ. Two mathematical objects shall be identical ($\equiv$) if all properties (observable and not observable) do not differ:

$$\delta(s) = \begin{cases} 1 & \text{for } s \equiv 0 \wedge s = 0 \\ 0 & \text{for } (s \not\equiv 0 \wedge s = 0) \vee (s \not\equiv 0 \wedge s \neq 0) \end{cases} \tag{1}$$

Here we introduce a realistic lack of precision. Because our resolution shall be limited, the interval fulfilling the condition $s \not\equiv 0 \wedge s = 0$ still contains an infinite number $n = \infty$ of locations s. There is only one location fulfilling the condition $s \equiv 0 \wedge s = 0$ versus said infinite number $n = \infty$ of locations with $s \not\equiv 0 \wedge s = 0$. Therefore, for a finite number $n \in \mathbb{N}$ of measurements, $\delta(s) = 0$ and $\delta(s) \not\equiv 0$ are valid. In the standard calculus, the anti-derivative of our point function $\delta(s)$ is zero because of the non-existing area between the point function $\delta(s)$ and the s-axis. We now cancel the principle of "natura non facit saltus" [22–25] to be able to handle the point functions $\delta(s)$. Thus, the whole mathematical standard calculus [26–35] is not applicable anymore. The standard calculus is incompatible with quantum mechanical calculations of point-like particles with no diameter [36, 37]. The subject of this paper is to overcome the inherent smoothness requirement conflict and to develop a hypothesis on the statistical origin of quantum mechanics.

2 The Divisible and Indivisible Problem, the Local vs. the Environmental Continuity and the Noise Function

In the first step, we introduce a summation on all elements s_i of the defnition set $\mathbb{V}$ of the unknown function $f(s)$ (symbol $\Sigma_{s_i \in \mathbb{V}}$) to solve the divisible and

indivisible problem [18–20]. If the summation is only performed on a limited number of $N \in \mathbb{N}$ (with $i \in \mathbb{N}$ and $1 \leq i \leq N$) known locations $s_i \in \mathbb{V}$, a point function sum remains equal but not identical to zero: $1 = \sum_{s_i \in \mathbb{V}} \delta(s - s_i) \wedge 0 = \sum_{i=1}^{N} \delta(s - s_i)$. A function $f(s)$ with the definition set $\mathbb{C}$ might be rewritten as:

$$f(s) = \sum_{s_i \in \mathbb{C}} f(s)\delta(s - s_i) = \sum_{s_i \in \mathbb{C}} f_i \delta(s - s_i) \text{ with } f(s_i) = f_i \tag{2}$$

A point function $\delta(s)$ might be seen as an imperfect new type of zero, which is equal but not identical to zero. This and the summation $\sum_{s_i \in \mathbb{V}}$ of such point functions $\delta(s - s_i)$ on all potential locations $s_i \in \mathbb{V}$ overcomes the divisible and indivisible problem. According to Cauchy [35] and Weierstraß, an unknown function $f(s) : \mathbb{C} \to \mathbb{C}$ is continuous if it fulfills $f(s) = \lim\limits_{\Delta s \to 0} f(s + \Delta s) = \lim\limits_{\Delta s \to 0} f(s - \Delta s)$. We define a weaker environmental continuity excluding the inspected location s and the function value $f(s)$ at the location s.

$$\lim_{\Delta s \to 0} f(s + \Delta s) = \lim_{\Delta s \to 0} f(s - \Delta s) \tag{3}$$

Therefore, a point function $\delta(s)$ is environmentally continuous but locally discontinuous. We define an environmental function of the function $f(s)$ as follows:

$$f_e(s) = \lim_{\Delta s \to 0} f(s + \Delta s) = \lim_{\Delta s \to 0} f(s - \Delta s) \tag{4}$$

The difference between the environment function value $f_e(s)$ and the function value $f(s)$ at the exact location s is a noise function $f_n(s)$.

$$f_n(s) = f(s) - f_e(s) = \sum_{i=1}^{N} \psi_i \delta(s - s_i) \tag{5}$$

Therefore, a hypothesis on the function value $f(s)$ based on the measurements should look like $f(s) = \lim\limits_{\Delta s \to 0} f(s + \Delta s) + \sum_{i=1}^{N} \psi_i \delta(s - s_i)$. The conditions $0 \not\equiv \sum_{i=1}^{N} \psi_i \delta(s - s_i) \wedge \sum_{i=1}^{N} \psi_i \delta(s - s_i) = 0$ are valid. Because we investigate unknown functions $f(s)$, N, the exact locations s_i of the point functions $\delta(s - s_i)$, and their corresponding amplitude ψ_i are unknown to us. Therefore, the description according to $f(s) = \lim\limits_{\Delta s \to 0} f(s + \Delta s) + \sum_{i=1}^{N} \psi_i \delta(s - s_i)$ is not correct because it is valid if we know the number N, the locations s_i, and the amplitude ψ_i. Therefore, we use formula (2) and change the amplitudes ψ_i into a locally continuous unknown distribution function $\psi(s) \not\equiv 0$ with the probability value $\psi(s_i) = \psi_i$ to take the unknown locations s_i into account in order to solve the continuum problem [38]. We define an environmental limit to reconstruct the noisy function $f(s)$ out of the conventional limit:

$$f(s) = \lim_{\Delta s \to 0} f(s + \Delta s) + \sum_{s_i \in \mathbb{C}} \psi(s)\delta(s - s_i) \tag{6}$$

3 Environmental Derivation

The differential quotient $\frac{\partial}{\partial_l s} f(s) = \lim\limits_{\Delta s \to 0} \frac{\pm f(s \pm \Delta s) \mp f(s)}{\Delta s}$ as defined by Cauchy [34] includes the location s and is therefore not applicable to point functions $\delta(s)$. We define an environmental derivation $\frac{\partial}{\partial_e s} f(s)$ based on the central differential quotient as used in finite element analysis [39]:

$$\frac{\partial}{\partial_e s} f(s) = \lim_{\Delta s \to 0} \frac{f(s + \Delta s) - f(s - \Delta s)}{2\Delta s} \tag{7}$$

Because the integral of a point function $\psi_i \delta(s - s_i)$ vanishes, the integration process deletes information about number N, amplitude ψ_i, and location s_i of the original point function portions in the initial function $f(s)$ before integration. To compensate for this information loss during a preceding integration, the deleted information must be reconstructed during differentiation. In combination with the environmental limes, as defined above in equation (6), we define the complete environmental derivation:

$$\frac{\partial}{\partial_{ec} s} f(s) = \frac{\partial f(s)}{\partial_e s} + \sum_{s_i \in \mathbb{C}} \psi(s) \delta(s - s_i) \tag{8}$$

with $\psi(s)$ being the probability to measure a point function $\delta(s - s_i)$ at the locations s_i. Our problem is that we do not know the original point function $\psi_i \delta(s - s_i)$ locations s_i, their amplitude ψ_i and their number N opreceding to integration. We do not know the amplitude $\psi(s_i)$ of the individual original point functions $\psi(s_i)\delta(s - s_i)$. Therefore, we might only assume that there is a probability $\psi(s)$ that there is a point function at the location s. All commonly known differentiation rules apply:

$$\begin{aligned} \frac{\partial h(s)}{\partial_{ec} s} &= \alpha \frac{\partial f(s)}{\partial_{ec} s} + \alpha \frac{\partial g(s)}{\partial_{ec} s} \\ \frac{\partial h(s)}{\partial_{ec} s} &= \frac{\partial f(s)}{\partial_{ec} s} \cdot g(s) + f(s) \cdot \frac{\partial g(s)}{\partial_{ec} s} \\ \frac{\partial h(s)}{\partial_{ec} s} &= \frac{\partial f(g)}{\partial_{ec} g} \cdot \frac{\partial g(s)}{\partial_{ec} s} \end{aligned} \tag{9}$$

4 Modified Quaternion Algebra and γ-Vectors

In the following sections, we use Einstein notation. We indicate space-like indexes, e.g. $I \in \{1, 2, 3\}$, by capital letters and other indexes, e.g. $i \in \{0, 1, 2, 3\}$ by lower case letters. The commonly used metrics in a flat space are $s^2 = x^2 + y^2 + z^2 - c^2 t^2$ with the distance s, the space like coordinates x, y, z, the velocity of light c, and time t. For simplification, we use natural coordinates with $c = 1$ and $\hbar = 1$. The length is then $s^2 = -x_0 x_0 + x_I x_I$ with x_0 being the time-like coordinate and x_I with $I \in \{1, 2, 3\}$ representing the 3

space-like coordinates. Usually [40], the metric is written using the metric tensor g_{ik} as $s^2 = g_{ik} \cdot x_i \cdot x_k$. We might write the $\boldsymbol{s}$ vector as $\boldsymbol{x} = \gamma_i \cdot x_i$. We name $\mathbb{V}$ the set of all possible γ-vectors $\boldsymbol{x} \in \mathbb{V}$. Correspondingly, we define the set $\mathbb{G} = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$ of Minkowski directions γ_i. The Hamiltonian quaternion concept [41] is known to cause some difficulties [42–48]. In contrast, we assume that the time-like direction γ_0 replaces the scalar component in the quaternion arithmetic. The arithmetic of the γ-vectors shall form an algebra with:

$$\begin{aligned} \alpha \cdot \boldsymbol{z} =& \gamma_i \cdot \alpha \cdot z_i = \gamma_i \cdot \alpha \cdot x_i + \gamma_i \cdot \alpha \cdot y_i = \alpha \cdot \boldsymbol{x} + \alpha \cdot \boldsymbol{y} \\ \alpha \cdot \boldsymbol{z} =& \boldsymbol{x} \cdot (\alpha \cdot \boldsymbol{y}) = (\boldsymbol{x} \cdot \alpha) \cdot \boldsymbol{y} \\ \boldsymbol{z} = \gamma_k \cdot z_k =& \gamma_k \cdot \lambda_{kij} \cdot x_i \cdot y_j = \gamma_i \cdot \gamma_j \cdot x_i \cdot y_j = \boldsymbol{x} \cdot \boldsymbol{y} \\ \boldsymbol{z}(\boldsymbol{x} + \boldsymbol{y}) =& \boldsymbol{z} \cdot \boldsymbol{x} + \boldsymbol{z} \cdot \boldsymbol{y} \end{aligned} \tag{10}$$

In contrast to the concept of Dirac [50] the set $\mathbb{G}$ of γ factors contains only 4 γ factors instead of 16 elements in Diracs's theory [51, 52]. The product rule of equation (10) in combination with $s^2 = g_{ik} x_i x_k$ and our approach that the γ_0 direction shall replace the quaternion scalar component let us define the square of $\gamma_s s = \gamma_i x_i$ as follows:

$$\boldsymbol{s}^2 = \gamma_0 \cdot s^2 = \gamma_0 \cdot g_{ij} \cdot x_i \cdot x_j = \gamma_i \gamma_j \cdot x_i \cdot x_j = \gamma_k \cdot \lambda_{kij} \cdot x_i \cdot x_j \tag{11}$$

Thus, we obtain the requirements:

$$g_{ij} = \lambda_{0ij} = \mathrm{diag}(-1, 1, 1, 1) \text{ and } \lambda_{kij} = -\lambda_{kji} \text{ for } i \neq j \tag{12}$$

We define a space-like portion $\boldsymbol{x}_s = \gamma_I \cdot x_I$ with $I \in \{1, 2, 3\}$ and a time-like portion $\boldsymbol{x}_t = \gamma_0 \cdot x_0$ of a γ-vector $\boldsymbol{x}$: $\boldsymbol{x} = \boldsymbol{x}_t + \boldsymbol{x}_s$. Condition (12) requests Clifford algebra [49]. For space-like $I, J, K \in \{1, 2, 3\}$ condition, (12) might be interpreted as a rotation in space with the rotation axis in γ_K direction. For the 3 space like γ-vector portions $\boldsymbol{x}_s, \boldsymbol{y}_s, \boldsymbol{z}_s \in \mathbb{V}$ we find:

$$\begin{aligned} \boldsymbol{z}_s =& \gamma_K \cdot z_K = \gamma_K \cdot \lambda_{KIJ} \cdot x_i \cdot y_j \\ =& \gamma_1 (x_2 \cdot y_3 - x_3 \cdot y_2) + \gamma_2 (x_3 \cdot y_1 - x_1 \cdot y_3) + \gamma_3 (x_1 \cdot y_2 - x_2 \cdot y_1) \end{aligned} \tag{13}$$

The most significant problem occurs for mixed time/space like portions. Out of the Clifford algebra condition (12), we obtain for those elements $\lambda_{k0J} = -\lambda_{kJ0}$ for $J \in \{1, 2, 3\}$. Together with the condition that the time-like γ_0 component should behave similar to a scalar, this corresponds to the $\gamma_J = \gamma_0 \gamma_J = -\gamma_J \gamma_0$ product condition for the γ-factors with $J \in \{1, 2, 3\}$. If the γ factors are matrices like the γ_μ matrices in the Dirac algebra, they are not able to fulfill condition $\gamma_J = \gamma_0 \gamma_J = -\gamma_J \gamma_0$. There is no non commutative unit matrix. The three

aspects of a sample version of the λ_{koj} matrix are:

$$\lambda_{0ij} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}, \lambda_{1ij} = \begin{pmatrix} 0 & +1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & +1 & 0 \end{pmatrix},$$
$$\lambda_{2ij} = \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \lambda_{3ij} = \begin{pmatrix} 0 & 0 & 0 & +1 \\ 0 & 0 & -1 & 0 \\ 0 & +1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_{i0j} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}, \lambda_{i1j} = \begin{pmatrix} 0 & +1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 \\ 0 & 0 & -1 & 0 \end{pmatrix},$$
$$\lambda_{i2j} = \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \end{pmatrix}, \lambda_{i3j} = \begin{pmatrix} 0 & 0 & 0 & +1 \\ 0 & 0 & +1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \tag{14}$$

$$\lambda_{ij0} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \lambda_{ij1} = \begin{pmatrix} 0 & +1 & 0 & 0 \\ +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & +1 & 0 \end{pmatrix},$$
$$\lambda_{ij2} = \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \\ +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \lambda_{ij3} = \begin{pmatrix} 0 & 0 & 0 & +1 \\ 0 & 0 & -1 & 0 \\ 0 & +1 & 0 & 0 \\ +1 & 0 & 0 & 0 \end{pmatrix}$$

This λ_{ijk}-matrix (14) has for $i \in \{0,1,2,3\}$, $j \in \{0,1,2,3\}$, $k \in \{0,1,2,3\}$, $I \in \{1,2,3\}$ and $K \in \{1,2,3\}$ and E as the unit matrix the properties

$$\begin{gathered} \lambda_{0ik} = g_{ik}\ ,\ \lambda_{0jk} = \lambda_{0kj}\ ,\ \lambda_{Ijk} = -\lambda_{Ikj}\ ,\ \lambda_{i0k} = g_{ik}\ ,\ \lambda_{i0k} = \lambda_{k0i} \\ \lambda_{iJk} = -\lambda_{kJi}\ ,\ \lambda_{ij0} = -E\ ,\ \lambda_{ij0} = \lambda_{ij0}\ ,\ \lambda_{0jK} = \lambda_{j0K}\ ,\ \lambda_{IJK} = -\lambda_{JIK} \end{gathered} \tag{15}$$

5 Complete Environmental Derivation of a γ-Vector Function $\boldsymbol{f(x)}$

Now, we compute the complete environmental derivation of a γ-vector function $\boldsymbol{f(x)}$. Because the product of two γ vectors is not commutative, we might distinguish between the left and the right handed derivation we discuss here only:

$$\begin{aligned} &\boldsymbol{f(x+\Delta x)} - \boldsymbol{f(x-\Delta x)} = \{\gamma_i f_i(\boldsymbol{x} + h\cdot\boldsymbol{\Delta x}) - \gamma_i f_i(\boldsymbol{x} - h\cdot\boldsymbol{\Delta x})\} \\ =&\gamma_i\lambda_{iml}\lambda_{jmk}\frac{\partial f_j(\boldsymbol{x})}{\partial_{ec}x_k}2\cdot h\cdot\Delta x_l - \gamma_i\lambda_{iml}\lambda_{pmq}e_{pq}(h\cdot\Delta x_l)\} \end{aligned} \tag{16}$$

By comparison of coefficients, division by $2h$, and computing the environmental limit, we obtain:

$$\begin{aligned}&\lim_{h\to 0}\frac{f_i(\boldsymbol{x}+h\cdot\boldsymbol{\Delta x})-f_i(\boldsymbol{x}-h\cdot\boldsymbol{\Delta x})}{2h}\\ =&\lim_{h\to 0}\lambda_{iml}\lambda_{jmk}\frac{\partial f_j(\boldsymbol{x})}{\partial_{ec}x_k}\cdot\Delta x_l-\lim_{h\to 0}\lambda_{iml}\lambda_{pmq}\frac{1}{2h}e_{lpq}(h\cdot\boldsymbol{\Delta x})\end{aligned}\tag{17}$$

The right-hand term vanishes by definition.

$$\lim_{h\to 0}\frac{f_i(\boldsymbol{x}+h\cdot\boldsymbol{\Delta x})-f_i(\boldsymbol{x}-h\cdot\boldsymbol{\Delta x})}{2h}=\lim_{h\to 0}\lambda_{iml}\lambda_{jmk}\frac{\partial f_j(\boldsymbol{x})}{\partial_{ec}x_k}\cdot\Delta x_l\tag{18}$$

For each individual direction $\gamma_l\Delta x_l$ and each intermediate direction γ_m, we find because of $\lambda_{iml}\cdot\lambda_{iml}=1$:

$$\lim_{h\to 0}\lambda_{iml}\frac{f_i(\boldsymbol{x}+h\cdot\boldsymbol{\Delta x})-f_i(\boldsymbol{x}-h\cdot\boldsymbol{\Delta x})}{2h\cdot\Delta x_l}=\lim_{h\to 0}\lambda_{jmk}\frac{\partial f_j(\boldsymbol{x})}{\partial_{ec}x_k}\tag{19}$$

If we remove the limit on the right side, we have to add the corresponding noise function $\sum_{\boldsymbol{x}_r\in\mathbb{V}}\lambda_{pqm}\mathfrak{A}_{pq}(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{x}_r)$:

$$\lambda_{iml}\lim_{h\to 0}\frac{f_i(\boldsymbol{x}+h\boldsymbol{\Delta x})-f_i(\boldsymbol{x}-h\boldsymbol{\Delta x})}{2h\cdot\Delta x_l}+\sum_{\boldsymbol{x}_r\in\mathbb{V}}\lambda_{pmq}\mathfrak{A}_{pq}(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{x}_r)=\lambda_{jmk}\frac{\partial f_j(\boldsymbol{x})}{\partial_{ec}x_k}\tag{20}$$

Because of $\sum_{\boldsymbol{x}_r\in\mathbb{V}}\delta(\boldsymbol{x}-\boldsymbol{x}_r)=1$ and by multiplication with γ_m, we find :

$$\gamma_m\lambda_{jmk}\frac{\partial f_j(\boldsymbol{x})}{\partial_e x_k}+\gamma_m\lambda_{pmq}\mathfrak{A}_{pq}(\boldsymbol{x})=\gamma_m\lambda_{jmk}\frac{\delta f_j(\boldsymbol{x})}{\partial_{ec}x_k}\tag{21}$$

6 Deduction of the Schrödiger-Dirac-Equation for Point Functions based on the Complete Environmental Derivation of $\boldsymbol{\gamma}$-Vectors

We start with the complete environmental constant effect function $\gamma_0 S(\boldsymbol{x})=\gamma_0\cdot S$ and with $\lambda_{pmq}\mathfrak{P}_{pq}(\boldsymbol{x})=\psi_m(\boldsymbol{x})$:

$$\gamma_m\lambda_{0mk}\frac{\partial S}{\partial_{ec}x_k}=\gamma_m\lambda_{pmq}\mathfrak{P}_{pq}(\boldsymbol{x})=\gamma_m\psi_m(\boldsymbol{x})\tag{22}$$

We apply the complete environmental derivation a second time, and obtain the following by comparison of coefficients:

$$\gamma_m\lambda_{jmk}\frac{\partial\psi_j(\boldsymbol{x})}{\partial_{ec}x_k}=\gamma_m\lambda_{jmk}\frac{\partial\psi_j(\boldsymbol{x})}{\partial_e x_k}+\gamma_m\lambda_{pmq}\mathfrak{A}^*_{qp}(\boldsymbol{x})\tag{23}$$

$\gamma_m\lambda_{pmq}\mathfrak{A}^*_{qp}(\boldsymbol{x})$ is the noise function. The point function space $\mathbb{V}$ is homogeneous:

$$0=\gamma_m\lambda_{jmk}\frac{\partial\psi_j(\boldsymbol{x})}{\partial_e x_k}+\gamma_m\lambda_{pmq}\mathfrak{A}^*_{qp}(\boldsymbol{x})\tag{24}$$

This homogeneity assumption corresponds to "natura non facit saltus" and the cosmological assumption of a homogeneous universe. The noise function $\mathfrak{A}_{qp}(\boldsymbol{x})$ is independent in its value and its indexes p,q of the original derived function. Its dimension indexes p, q might be scrambled vs. the other indexes j, k. Renaming without documenting this scrambling might delete this difference. We therefore multiply each coordinate with a flipping matrix ${}^{u}\Xi_{kq}$ or ${}^{u}\Xi_{jp}$ to connect the other indexes. The index u stands for the type of scrambling:

$$0 = \gamma_m \lambda_{jmk} \frac{\partial \psi_j(\boldsymbol{x})}{\partial_e x_k} + \gamma_m \lambda_{jmk} \ {}^{u}\Xi_{kq} \ {}^{u}\Xi_{jp} \mathfrak{A}^*_{qp}(\boldsymbol{x}) \tag{25}$$

There are $4! = 24$ values of u. We define the u-index specific vector potential density ${}^{u}\mathfrak{A}_{qj}(\boldsymbol{x})$, the u-index specific vector potential ${}^{u}A_q(\boldsymbol{x})$ and the density distribution function $\psi_j(\boldsymbol{x})$ as follows:

$${}^{u}A_q(\boldsymbol{x}) \cdot \psi_j(\boldsymbol{x}) = \ {}^{u}\mathfrak{A}_{qj}(\boldsymbol{x}) = \ {}^{u}\Xi_{jp} \mathfrak{A}^*_{qp}(\boldsymbol{x}) \tag{26}$$

Thus, we obtain:

$$0 = \gamma_m \lambda_{jmk} \frac{\partial \psi_j(\boldsymbol{x})}{\partial_e x_k} + \sum_{u=0}^{23} \gamma_m \lambda_{jmk} \ {}^{u}\Xi_{kq} \ {}^{u}A_q(\boldsymbol{x}) \cdot \psi_j(\boldsymbol{x}) \tag{27}$$

6.1 Stable Solutions, Mass Gap an Point Function Spin

We define a point function $\delta(\boldsymbol{x} - \boldsymbol{x}_i)$ to be stable if the divergence of the corresponding density distribution function $\boldsymbol{\psi(x)}$ vanishes.

$$0 = \gamma_0 \lambda_{j0k} \frac{\partial \psi_j(\boldsymbol{x})}{\partial_e x_k} = \gamma_0 \lambda_{0jk} \frac{\partial \psi_j(\boldsymbol{x})}{\partial_e x_k} \tag{28}$$

From (27) with $\lambda_{j0k} = \lambda_{0jk}$, we find for the time-like portion:

$$0 = \gamma_0 \lambda_{j0k} \frac{\partial \psi_j(\boldsymbol{x})}{\partial_e x_k} + \sum_{u=0}^{23} \gamma_0 \lambda_{j0k} \ {}^{u}\Xi_{kq} \ {}^{u}A_q(\boldsymbol{x}) \cdot \psi_j(\boldsymbol{x}) \tag{29}$$

The mass $\boldsymbol{m}(\boldsymbol{x})$ shall be $-m_j(\boldsymbol{x}) = \sum_{u=0}^{23} \lambda_{0jk} \ {}^{u}\Xi_{kq} \ {}^{u}A_q(\boldsymbol{x})$.

$$0 = \gamma_0 \lambda_{j0k} \frac{\partial \psi_j(\boldsymbol{x})}{\partial_e x_k} - \gamma_0 m_j(\boldsymbol{x}) \cdot \psi_j(\boldsymbol{x}) \tag{30}$$

To limit equation (27) to stable solutions fulfilling condition (28), we subtract equation (30) from equation (27). The resulting equation (31) is space-like because m=0 is deleted.

$$0 = \gamma_M \lambda_{jMk} \frac{\partial \psi_j(\boldsymbol{x})}{\partial_e x_k} + \sum_{u=0}^{23} \gamma_M \lambda_{jMk} \ {}^{u}\Xi_{kq} \ {}^{u}A_q(\boldsymbol{x}) \cdot \psi_j(\boldsymbol{x}) + \gamma_0 m_j(\boldsymbol{x}) \cdot \psi_j(\boldsymbol{x}) \tag{31}$$

To solve equation (31), we might multiply (31) with $\frac{\gamma_M \pm \gamma_0}{2}$ from the left:

$$\begin{aligned} 0 =& \frac{1}{2}\{\gamma_M\gamma_M\lambda_{jMk}\frac{\partial\psi_j(\boldsymbol{x})}{\partial_e x_k} + \sum_{u=0}^{23}\gamma_M\gamma_M\lambda_{jMk}\ {}^u\Xi_{kq}\ {}^uA_q(\boldsymbol{x})\cdot\psi_j(\boldsymbol{x}) + \gamma_M\gamma_0 m_j(\boldsymbol{x})\cdot\psi_j(\boldsymbol{x})\} \\ \pm& \frac{1}{2}\{\gamma_0\gamma_M\lambda_{jMk}\frac{\partial\psi_j(\boldsymbol{x})}{\partial_e x_k} + \sum_{u=0}^{23}\gamma_0\gamma_M\lambda_{jMk}\ {}^u\Xi_{kq}\ {}^uA_q(\boldsymbol{x})\cdot\psi_j(\boldsymbol{x}) + \gamma_0\gamma_0 m_j(\boldsymbol{x})\cdot\psi_j(\boldsymbol{x})\} \end{aligned} \tag{32}$$

With $\gamma_M\gamma_M = \gamma_0$, because of $m \in \{1,2,3\}$ and $\gamma_0\gamma_0 = -\gamma_0$ and $\gamma_0\gamma_M = \gamma_M = -\gamma_M\gamma_0$ and after resorting this is equal to:

$$0 = \frac{\gamma_M \pm \gamma_0}{2}\{\lambda_{jMk}[\frac{\partial\psi_j(\boldsymbol{x})}{\partial_e x_k} + \sum_{u=0}^{23}\ {}^u\Xi_{kq}\ {}^uA_q(\boldsymbol{x})\cdot\psi_j(\boldsymbol{x})] - m_j(\boldsymbol{x})\cdot\psi_j(\boldsymbol{x})\}\} \tag{33}$$

6.2 Potential Ξ-Matrices

To be able to investigate the properties of the Ξ-matrices, we will now write down all potential flipping operations.

$$\begin{aligned}
{}^0\Xi_{qm} &= \begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix}, \quad {}^1\Xi_{qm} = \begin{pmatrix} 0&1&0&0\\ 1&0&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix}, \quad {}^2\Xi_{qm} = \begin{pmatrix} 0&0&1&0\\ 0&1&0&0\\ 1&0&0&0\\ 0&0&0&1 \end{pmatrix} \\
{}^3\Xi_{qm} &= \begin{pmatrix} 0&0&0&1\\ 0&1&0&0\\ 0&0&1&0\\ 1&0&0&0 \end{pmatrix}, \quad {}^4\Xi_{qm} = \begin{pmatrix} 1&0&0&0\\ 0&0&1&0\\ 0&1&0&0\\ 0&0&0&1 \end{pmatrix}, \quad {}^5\Xi_{qm} = \begin{pmatrix} 1&0&0&0\\ 0&0&0&1\\ 0&0&1&0\\ 0&1&0&0 \end{pmatrix} \\
{}^6\Xi_{qm} &= \begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&0&1\\ 0&0&1&0 \end{pmatrix}, \quad {}^7\Xi_{qm} = \begin{pmatrix} 0&1&0&0\\ 1&0&0&0\\ 0&0&0&1\\ 0&0&1&0 \end{pmatrix}, \quad {}^8\Xi_{qm} = \begin{pmatrix} 0&0&1&0\\ 0&0&0&1\\ 1&0&0&0\\ 0&1&0&0 \end{pmatrix} \\
{}^9\Xi_{qm} &= \begin{pmatrix} 0&0&0&1\\ 0&0&1&0\\ 0&1&0&0\\ 1&0&0&0 \end{pmatrix}, \quad {}^{10}\Xi_{qm} = \begin{pmatrix} 0&1&0&0\\ 0&0&1&0\\ 1&0&0&0\\ 0&0&0&1 \end{pmatrix}, \quad {}^{11}\Xi_{qm} = \begin{pmatrix} 0&0&1&0\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&0&1 \end{pmatrix} \\
{}^{12}\Xi_{qm} &= \begin{pmatrix} 0&0&1&0\\ 0&1&0&0\\ 0&0&0&1\\ 1&0&0&0 \end{pmatrix}, \quad {}^{13}\Xi_{qm} = \begin{pmatrix} 0&0&0&1\\ 0&1&0&0\\ 1&0&0&0\\ 0&0&1&0 \end{pmatrix}, \quad {}^{14}\Xi_{qm} = \begin{pmatrix} 0&1&0&0\\ 0&0&0&1\\ 0&0&1&0\\ 1&0&0&0 \end{pmatrix}
\end{aligned} \tag{34}$$

$$
{}^{15}\Xi_{qm} = \begin{pmatrix} 0&0&0&1\\ 1&0&0&0\\ 0&0&1&0\\ 0&1&0&0 \end{pmatrix},\quad
{}^{16}\Xi_{qm} = \begin{pmatrix} 1&0&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 0&1&0&0 \end{pmatrix},\quad
{}^{17}\Xi_{qm} = \begin{pmatrix} 1&0&0&0\\ 0&0&0&1\\ 0&1&0&0\\ 0&0&1&0 \end{pmatrix}
$$

$$
{}^{18}\Xi_{qm} = \begin{pmatrix} 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 1&0&0&0 \end{pmatrix},\quad
{}^{19}\Xi_{qm} = \begin{pmatrix} 0&1&0&0\\ 0&0&0&1\\ 1&0&0&0\\ 0&0&1&0 \end{pmatrix},\quad
{}^{20}\Xi_{qm} = \begin{pmatrix} 0&0&1&0\\ 1&0&0&0\\ 0&0&0&1\\ 0&1&0&0 \end{pmatrix} \tag{35}
$$

$$
{}^{21}\Xi_{qm} = \begin{pmatrix} 0&0&0&1\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0 \end{pmatrix},\quad
{}^{22}\Xi_{qm} = \begin{pmatrix} 0&0&1&0\\ 0&0&0&1\\ 0&1&0&0\\ 1&0&0&0 \end{pmatrix},\quad
{}^{23}\Xi_{qm} = \begin{pmatrix} 0&0&0&1\\ 0&0&1&0\\ 1&0&0&0\\ 0&1&0&0 \end{pmatrix}
$$

The Ξ matrices are sorted as follows: 1^{st} group: ${}^{A}\Xi$ matrix, with ${}^{A}\Xi = E_{qm}{:}^{0}\Xi$; 2^{nd} group: 6 ${}^{B}\Xi$ matrices, with ${}^{B}\Xi^2 = E_{qm}{:}^{1}\Xi$ to ${}^{6}\Xi$; 3^{rd} group: 3 ${}^{C}\Xi$ matrices, with ${}^{C}\Xi^2 = E_{qm}{:}^{7}\Xi$ to ${}^{9}\Xi$; 4^{th} group: 8 ${}^{D}\Xi$ matrices, with ${}^{D}\Xi^3 = E_{qm}{:}^{10}\Xi$ to ${}^{17}\Xi$; 5^{th} group: 6 ${}^{E}\Xi$ matrices, with ${}^{E}\Xi^4 = E_{qm}{:}^{18}\Xi$ to ${}^{23}\Xi$.

6.3 Velocity and Metrics

We assume now a recursive coordinate definition $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{V}$. The interdependence between the coordinates x_i, x_j modifies equation (33):

$$
0 = \frac{\gamma_M \pm \gamma_0}{2}\{\lambda_{jMk}[\frac{\partial\psi_j(\boldsymbol{x})}{\partial_e x_r}\frac{\partial x_r}{\partial_e x_k} + \sum_{u=0}^{23} {}^{u}\Xi^{u}_{kq} A_q(\boldsymbol{x})\psi_j(\boldsymbol{x})] - m_j(\boldsymbol{x})\psi_j(\boldsymbol{x})\} \tag{36}
$$

We multiply equation (36) with the velocity $v_{rk} = \frac{\partial x_k}{\partial_e x_r}$ with $v_{rk} \cdot \frac{\partial x_r}{\partial_e x_k} = E_{rr}$

$$
0 = \frac{\gamma_M \pm \gamma_0}{2}\{\lambda_{jMk}[\frac{\partial\psi_j(\boldsymbol{x})}{\partial_e x_k} + \sum_{u=0}^{23} v_{rk}\ {}^{u}\Xi^{u}_{rq} A_q(\boldsymbol{x})\psi_j(\boldsymbol{x})] - v_{jr} m_r(\boldsymbol{x})\psi_j(\boldsymbol{x})\} \tag{37}
$$

7 Interpretation, Discussion and Outlook

7.1 Point function particle Interpretation

Equation (37) is similar to the relativistic Dirac-Schrödinger equation [51, 53] for fermionic elementary particles. $\gamma^\mu(i\hbar\partial_\mu\psi - \frac{e}{c}A_\mu\psi) - m^*c\psi = 0$ To convert equation (37) into this equation we eliminate the factor $\frac{\gamma_m \pm \gamma_0}{2}$ in equation (37) by comparison of coefficients. We reduce equation (37) to one dedicated index j (means $\psi_j(\boldsymbol{x}) = \psi$) and multiply equation (37) with the constant $i \cdot \hbar$. We define $i \cdot \hbar \cdot c \cdot\ {}^{u}\Xi_{kq} = e$ and $i \cdot \hbar \cdot \frac{m_j(\boldsymbol{x})}{c} = m^*$ and $\frac{\partial}{\partial_e x_k} = \partial_\mu$ and obtain for $v_{rk} = E_{rk}$ the equation $\gamma^\mu(i\hbar\partial_\mu\psi - \frac{e}{c}A_\mu\psi) - m^*c\psi = 0$. The spin γ_μ corresponds to λ_{jMk}.

If we investigate equation (37) for each u, we might separate one u index specific equation:

$$0 = \lambda_{jMk}[\frac{\partial \psi_j(\boldsymbol{x})}{\partial_e x_k} + v_{rk}\ {}^u\Xi^u_{rq} A_q(\boldsymbol{x})\psi_j(\boldsymbol{x})] - v_{jr} m_r(\boldsymbol{x})\psi_j(\boldsymbol{x}) \tag{38}$$

With respect to equation (14) and because of $M \in \{1, 2, 3\}$ due to the limitation on stable solutions, we recognize that there are always two solutions of equation (38). They represent a point function particle and its point function anti-particle. Because there are 24 possibilities for the index $u \in \{0, 1, ...22, 23\}$, the number is increased to $2 \cdot 24 = 48$ point function particles. The term $\frac{\gamma_M \pm \gamma_0}{2}$ in equation (37) allows two spin orientations: A) spin up (+) and B) spin down (−).

7.2 Organization of Point Function Particles

The point function particles might be organized depending on the coordinate scrambling. There are 4 initial possibilities to map the time-like direction γ_0 to a) 1 time-like direction γ_0 (time-like point function particles) and b) 3 space-like directions γ_1, γ_2, γ_3 (space-like point function particles). Each of these four point function particle groups contains 6 point function particles.
Time-Like Point Function Particles: The first 6 time-like point function particles might be resorted depending on the coordinate that the time-like direction γ_0 is mapped on. In the case of a time-like point function particle, there are 3 possibilities left to map the time-like direction γ_0 of the remaining 3 space-like directions γ_1, γ_2, γ_3 to. For each of these sub-classes, there are two possibilities of how to handle the remaining two coordinates.
Space Like Point Function Particles: The 3 groups of 6 space-like point function particles might be resorted depending on the direction the corresponding space-like direction γ_I with $I \in \{1, 2, 3\}$ is mapped on. In this case there are 3 possible directions λ_0 and the other two space-like directions λ_J, λ_K with $J \neq I$ and $J \neq K$ and $I \neq K$ left to map the space-like direction γ_I to. For each of these sub-classes, there are again two possibilities of how to handle the remaining two coordinates.
γ-Factor caused Structure of Point Function Particles: The structure is incorporated in the γ-factors and the matrices λ_{pMq} and λ_{iMk} of equation (24) as well. There are four pairs of indexes j, k. One pair might be named time-like if $i = 0$. For stability reasons, as pointed out above $m = 0$ is excluded. The terms with λ_{p0q} and λ_{i0k} are subtracted when forming the mass term. Therefore, 3 space-like possibilities for $M \in \{1, 2, 3\}$ are and 4 pairs of field oriented indexes p, q are left. If we investigate the space-like γ-factors γ_1, γ_2, and γ_3, we recognize that for all of them, half of the coordinate transformation factors are positive $+1$ and the other half negative -1. As known from the solution of Dirac equation, this causes the two-point function particle/point function anti-particle solution to appear.

7.3 Conclusion and Outlook

As shown above, it is possible to construct a new type of calculus allowing the differentiation of point functions (1) marking events and particles and Dirac's delta function. The inner core of the new calculus consists of a) the central differential quotient (7), named environmental derivation, b) the environmental continuity (3), c) the complete limit (6), d) the complete environmental derivation (8). The fundamental theorem of analysis [54] for unnknown functions $f(s)$ transforms to:

$$\int \frac{\partial f(s)}{\partial_{ec} ds} ds = f(s) + c + \sum_{s_i \in \mathbb{V}} (\frac{\partial \psi(s)}{\partial_e s} + \Phi(s)) \delta(s - s_i)$$

The application to point function wave equations has been demonstrated. Further research on the application to the statistics of unknown non-continuous functions should be performed. Further research is required to determine if the similarities reported above are occasional similarities or if they have a real physical background. Furthermore, more research on the modified quaternion concept and the concept of complete environmental calculus is required to prove and confirm all the above statements from all potential aspects. The similarities to the standard model must be subject of a future investigations and publications.

8 Acknowledgment

The author thanks J. Meijer, P. Glöseкötter, W. R. Fahrner, B. Steinsiepe, Benedikt Burchard, M. Burchard and R. Verch for critical support.

References

1. Falkenburg, B.: Particle Metaphysiscs A Critical Account of Subatomic Reality, Springer, Berlin (2007)
2. Bohr, N.: On the Constitution of Atoms and Molecules, Part II Systems Containing Only a Single Nucleus, Philosophical Magazine **26** (153), 476502,(1913)
3. Bohr, N.: On the Constitution of Atoms and Molecules, Part I, Philosophical Magazine **26** (151), 124 (1913)
4. Dirac, P.: The Principles of Quantum Mechanics, 4th edn., Oxford at the Clarendon Press, (1958)
5. Duistermaat, J. J., Kolk J. A. C. Distributions: Theory and Applications (Cornerstones), Springer, New York (2010)
6. Schwartz, L., Théorie des distributions, SERBIULA (sistema Librum 2.0) (2018)
7. Colombeau, J. F., Multiplication of Distributions: A tool in mathematics, numerical engineering and theoretical physics, Springer, New York, (1992)
8. Guillemin, V., Some Micro-Local Aspects of Analysis on Compact Symmetric Spaces, Princeton University Press, Princeton (1979)
9. Vasy, A., Microlocal analysis of asymptotically hyperbolic and Kerr-de Sitter spaces, Invent. math. **194**, 2, 381 (2013)
10. Kaneko, A., Introduction to Hyperfunctions, Springer, New York (1988)
11. Hörmander, L., The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis, Springer, New York (2013)

12. Robinson, A.: Non-standard Analysis, American Elsevier Pub. Co. New York, (1974)
13. Bell J. L.: A Primer if infinitesimal Analysis, 2^{nd} edn. Cambridge University Press, New York (2008)
14. Moerdijk I., Reyes G. E., Models for Smooth Infinitesimal Analysis, Springer, New York, (1991)
15. Wakabayashi, S., Classical Microlocal Analysis in the Space of Hyperfunctions, Springer (2000)
16. Ali, S.T., Antoine, J-P., Gazeau, J-P. Coherent States, Wavelets and Their Generalizations: A Mathematical Overview, Springer, New York (2000)
17. Seuret, S., Multifractal Analysis and Wavelets. New Trends in Applied Harmonic Analysis, HAL Id: hal-01612275, ttps://hal.archives-ouvertes.fr/hal-01612275 (2016), last download: 27.06.2018
18. Cassirer, E., Ed.Recki, B., Simn, M.: Bd. 1: Leibniz' System in seinen wissenschaftlichen Grundlagen, Meiner, Hamburg (1998)
19. Guénon, R.: The Metaphysical Principles of the Infinitesimal Calculus, Sophia Perennis, New York (2003)
20. Guénon,R., Les Principes du Calcul Infinitésimal, Éditions Gallimard, (1946)
21. Bondar, D. I., Lompay, R.R., Liu, W.K.: Quantum mechanics of a free particle from properties of the Dirac delta function,arXiv:1007.4243v2 [math-ph] (18 Mar 2011)
22. Linné, C. v.: Philosophia Botanica. Stockholm 1751, Ed 2003, Oxford University Press, New York (2003)
23. Leibniz, G. W.: Neue Abhandlungen über den menschlichen Verstand, (1704) Philosophische Bibliothek (Book 498), Felix Meiner, Berlin (2017)
24. Comenius, J. A.: De sermonis Latini studio, Bayerische Staatsbibliothek, Müller, (1638)
25. Marshall A.:1890, Principles of Economics. Abridged Ed., Cosimo, New York (2009)
26. Descartes, R.: 1637/1669, La Gomtrie. In Ed. Hermann A., La Gomtrie, Ste Exploitation Domain (2007)
27. Van Schooten, F.: 1683, Geometria A Renato Descartes Anno 1637 Gallice Edita. Latin Edition, Kessinger Publishing, LLC (2010)
28. Newton. I., 1665, A method whereby to square those crooked lines which may bee square. In Mathematical Papers, **1**, 302-313, Cambridge University Press, New York (2008)
29. Newton, I.: 1669, De analysi per aequationes infinitas. In Mathematical Papers, **2**, 206-247, Cambridge University Press, New York (2008)
30. Newton, I.: 1670-1671, De mehodis serierum et fluxionum. In Mathematical Papers, **3**; 32-329, Cambridge University Press, New York (2008)
31. Newton, I.: 1687/1963, Philosohiae Naturalis Principia Mathematica, London. In German Ed., Wolfers J. Ph., Wissenschaftliche Buchgesellschaft, Darmstadt (reprint of Berlin edition 1872)
32. Leibnitz, G. W.: 1684, Nova Methodus pro maximis et minimis, itemque tangentibus, quae nec fractas nec irrationales quantitates moratur et singulare pro illis calculi genus. Acta Eruditorum, 467-473 (1684)
33. Leibnitz, G. W.: 1686, De geometria recondita et analysi invisibilium atque infinitorum. Acta Eruditorum, 192 (1686)
34. Cauchy, A.L.: Résumé des leçons données à l'Ecole Polytechnique sur le calcul infinitśimal., Imprimerie Royale, Paris (1823)

35. Cauchy, A.L.: 1821/1885 Cours d'Analyse de Ecole Polytechnique. I^{me} partie. Analyse algébrique. Imprimerie Royale, Paris (1821)
36. Henley, E. M., Garcia A., Chapter 6.5 Leptons are Point Particles. In Subatomic Physics, 3^{rd} edn. 147, World Scientific Pub Co (2006)
37. Dehmelt H. G.: Experiments with an isolated subatomic particle at rest, Nobel Lecture, Department of Physics, University of Washington, Seattle, 590 (1989)
38. Taschner, R.: The Continuum, Vieweg (2005)
39. Leismann H. M. , Herrling B.: Two-dimensional finite element schemes for an improved computation of convective transport in fluids. In Hirschel, H.: Finite Approximations in Fluid Mechanics II: DFG Priority Research Programme Results 1986-1988; 264-278 (1989)
40. Moeller, C.: The theory of Relativity, Oxford University Press, London (1972)
41. Rosenfeld B. A.: A History of Non-Euclidean Geometry, Springer, New York (1988)
42. Rawat, A. S., Rawat S., Negi O.P.S.: Quaterinonic Formulation of Dirac Equation, IJTAS **3** (2) 1-5 (2011)
43. Welch, L.C.: Dirac's Equation in Different Numerical Rings, arXiv:0806.3396 [physics.gen-ph] (3.2.2009)
44. Welch, L. C.: A Possible Mathematical Structure for Physics,arXiv:0908.2063v1 [physics.gen-ph] (14.8.2009)
45. Hamich, S., Abbas, H: Two dimensional representation of the Dirac equation in Non associative algebra, arXiv:1104.3416v1 [math-ph] (18.04.2011)
46. Antonuccio, F.: Split-Quaternions and the Dirac Equation, arXiv:1405.0187v4 [physics.gen-ph] (7.7.2014)
47. Rowlands, P., Cullerne, J. P.: The Dirac Algebra and its Physical Interpretation, arXiv:quant-ph/0010094 (27.10.2000)
48. De Leo, S., Rodridues, W.: Quaternionic Electron Theory: Dirac's Equation, Int. J. Theor. Phys. **37**(5) 15111529 (1998)
49. Hladik, J.: Spinors in Physics, New York (1999)
50. Macfarlane, A. J.: Dirac Matrices and the Dirac Matrix Description of Lorentz Transformations, Commun. mth. Phys. **2** 133-146 (1966)
51. Dirac, P. A. M.: Die Prinzipien der Quantenmechanik, Hirzel, Leipzig, Germany (1930)
52. Kim, Y. S., Noz, M. E.: Feynmans Rest of the Universe, arXiv:1210.6251, (23.10.2012)
53. Rebhan, E.: Theroretische Physik Band 2, Elsevier, Mnchen (2005), 865
54. Stroyan, K. D., Luxemburg W. A. J.: Introduction to the Theory of Initesimals, Academic Press New York (1076)

Dynamic Bayesian smooth transition autoregressive models applied to hourly electricity load in southern Brazil

Álvaro E. Faria
and
Alexandre J. Santos

July 19, 2018

Abstract

Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models are proposed for nonlinear autoregressive time series processes as alternative to both the classical Smooth Transition Autoregressive (STAR) models of Chan and Tong (1986) and the Bayesian Simulation STAR (BSTAR) models of Lopes and Salazar (2005). Unlike those, DBSTAR models are sequential polynomial dynamic analytical models suitable for inherently non-stationary time series with non-linear characteristics such as asymmetric cycles. As they are analytical, they also avoid potential computational problems associated with BSTAR models and allow fast sequential estimation of parameters.

Two types of DBSTAR models are defined here based on the method adopted to approximate the transition function of their autoregressive components, namely the Taylor and the B-splines DBSTAR models. A harmonic version of those models, that accounted for the cyclical component explicitly in a flexible yet parsimonious way, were applied to the well-known series of annual Canadian lynx trappings and showed improved fitting when compared to both the classical STAR and the BSTAR models. Another application to a long series of hourly electricity loading in southern Brazil, covering the period of the South-African Football World Cup in June 2010, illustrates the short-term forecasting accuracy of fast computing harmonic DBSTAR models that account for various characteristics such as periodic behaviour (both within-the-day and within-the-week) and average temperature.

Keywords: Bayesian dynamic STAR models, polynomial forecasting models, nonlinear autoregressive models, Bayesian autoregressive forecasting models, short-term electricity load forecasting, B-splines approximation.

School of Mathematics and Statistics, Faculty of Science Technology, Engineering and Mathematics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK.

1 Introduction

The proposed DBSTAR models consist of Gaussian Bayesian state-space formulations based on polynomial *dynamic linear models* (DLMs) of West and Harrison (1997). They extend *smooth transition autoregressive* (STAR) models of Chan and Tong (1986) to allow parameters such as autoregressive, smoothing and observational variance to change in time. Parametric assessment is analytical with sequential prior-to-posterior distributional updating carried out by Kalman filtering for fast computation.

STAR models were developed from the *threshold autoregressive* (TAR) models proposed by Tong (1978) to address the failures of single linear models to represent certain properties of nonlinear stationary autoregressive time series processes such as asymmetric cycles, amplitude dependent frequencies and sudden changes (see Tong, 2011 for a review of developments of TAR type models over the last 30 years).

Basically, a STAR (and a TAR) model can be seen as a convex linear combination of two (or more) distinct linear *autoregressive* (AR) models of the same order. Each of the models is given a weight $w_i \in [0,1]$ $(i = 1, 2)$ in the combination that will average out the models. As the combination is convex, $\sum_i w_i = 1$, when a weight w_i is either 0 or 1, one of the combining AR models will be selected to operate. For that reason, a TAR or a STAR model is called a *regime-switching* model. Usually, the combining weights are specified through a conveniently chosen function which type defines the type of model. For instance, if a logistic function is adopted, the model is called a logistic STAR model.

For a stationary time series Y_t $(t = 1, 2, \ldots, T)$, a Gaussian STAR model of order p, STAR(p), $p \in \Re^+$, with two *regimes*, can be represented by the combination

$$Y_t = \pi(\cdot)\underline{z}_t\underline{\phi}_1' + [1 - \pi(\cdot)]\underline{z}_t\underline{\phi}_2' + \epsilon_t \quad ; \quad \epsilon_t \sim N(0, \sigma^2) \tag{1}$$

where for $i = 1, 2$, $\underline{\phi}_i = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{ip})$ are $(p+1)$-dimensional vectors with element ϕ_{ij} $(j = 0, 1, \ldots, p)$ representing an AR coefficient associated with each component j of the regime i; $\underline{z}_t = (1, y_{t-1}, \ldots, y_{t-p})$ is a $(p+1)$-dimensional vector with element y_{t-j} representing a realisation of the process Y_{t-j} at time $t - j$. The weight $\pi(\cdot)$, called transition function, is a function (of its arguments only) in the range [0, 1]. Note that in this paper, an underlined character is used to represent a vector, a matrix is represented by a boldface capital character and a prime is used to denote transposition.

The STAR models of Chan and Tong (1986) use conveniently chosen smooth transition functions, such as the logistic function

$$\pi(s_t; \gamma, c) = [1 + \exp\{-\gamma(s_t - c)\}]^{-1}, \tag{2}$$

adopted here. It has smoothness and location parameters $\gamma \in \Re^+$ and $c \in \Re$, respectively, and transition variable $s_t \in \Re$. Usually, in practice, the transition variable is either a a lagged past value y_{t-d}, where d is a delay parameter, or a chosen exogenous variable. The parameter γ dictates the degree of smoothness of $\pi(s_t; \gamma, c)$ and c is a threshold value between the two regimes. For the same value of γ, the distance between the value of s_t and c determines the degree of pertinence between the two regimes. For values of γ leaning towards zero the logistic function tends to $1/2$ and the *logistic* STAR model is reduced to the average between the combining AR models. As γ increases away from zero,

the logistic function tends to a step function and the transition from one regime to the other becomes more abrupt. Note that during a transition period, the combining model is non-linear in form and usually non-linear least square approaches, that approximate the non-linear transition function, are adopted for parametric estimation. We refer to those STAR models as *classical* STAR models throughout.

Bayesian approaches for TAR models, and their variants, were initially proposed by Geweke and Terui (1993) and Chen and Lee (1995), and further developed by Chen (1998), Campbell (2004), Lubrano (2000) and Lopes and Salazar (2005). All those approaches are based on Markov Chain Monte Carlo (MCMC) simulation methods, such as the sampling importance resampling of Gelfand and Smith (1990), for Bayesian inference due to the loss of analytical tractability of posterior distributions as shown in Bauwens et al. (1999).

Except for Lopes and Salazar (2005), that treat the order p of the combining AR models as unknown and proceed to estimate it's value from data, all the other proposed methods assume p to be fixed a priori. In fact, Lopes and Salazar (2005) adopted a Gibbs sampler approach for inferences about $\underline{\phi}_1, \underline{\phi}_2, \gamma, c, d$ and σ^2 of the logistic STAR when p is considered known, and a reversible jump MCMC algorithm (Green, 1995) for posterior assessments when p is unknown. Thereafter, the models of Lopes and Salazar (2005) will be referred to as *Bayesian Simulation STAR* (BSTAR) models.

A common characteristic of all those approaches is that they are, without exception, static non-sequential methods for non-linear but stationary AR processes. It is also worth noting that computational Bayesian inference approaches, such as the BSTAR models, are non-parsimonious computer intensive numerical simulation models that rely on the availability of extensive data sets and on the possible convergence of chains to obtain approximate posterior distributions of underlying parameters. They are, consequently, not generally appropriate for applications that require fast sequential prior-to-posterior parametric estimation and forecasting.

The DBSTAR models address some of those limitations. Similarly to the classical STAR models, the AR order p and the delay parameter d are fixed a priori in a DBSTAR model. When those parameters are unknown, and initial data is available, a model selection approach can be adopted to determine their optimal values. Also, like BSTAR models, prior distributions must be specified for the state parameters of a DBSTAR model. Those parameters are functions of the AR coefficients and the smoothness parameter as well as the observational variance associated with STAR type models. However, unlike both the classical STAR and the BSTAR models, the observational variance in a DBSTAR model is not constant but allowed to vary in time, albeit slowly and steadily, to account for possible extra variation in the series. The slow changes in the observational variance of a DBSTAR model are determined sequentially from data in an approach based on the Kalman filter.

Another difference from both classical STAR and BSTAR models, is that a DBSTAR model can be parsimoniously formulated to explicitly account for components observed in the underlying time series, that is, level, trend, seasonality and cycle. A DBSTAR model is based on particular formulations of the polynomial *dynamic linear model* (DLM) of West and Harrison (1997) that allows hierarchical component modelling by superposition of models appropriate for specific components.

The main results of a comparative analysis against the classical and the Bayesian STAR models in the Canadian lynx application are shortly described here, while the application to the electricity load in southern Brazil is shown in more detail. More detailed analysis of both applications can be found in Santos (2014) and Faria and Santos (2018).

In both the Canadian lynx trapping and the Brazilian electricity loading applications, harmonic DBSTAR models fitted well the data. The Canadian Lynx data application, apart from showing the improved fitting performances for a harmonic DBSTAR model when compared with both the classical and the BSTAR models, also illustrated how the changing cyclic pattern, amongst others, can be dynamically estimated from the data. In the electricity load application, the formulated harmonic DBSTAR models showed considerably improved fitting over a SARIMA model obtained by SPSS's expert forecasting modeler. The B-splines slightly outperformed the Taylor harmonic DBSTAR model for both fitting and short-term 24-hours-ahead forecasting. The Taylor model performed slightly better for the longer 72-hours-ahead forecasting horizons.

This article is structured as follows. In Section 2, harmonic DBSTAR (HDBSTAR) models are formally defined for the Taylor and the B-splines approximations. Section 3 briefly describes the main results of a comparative fitting analysis between DBSTAR models and both the classical STAR and the BSTAR models when applied to the Canadian lynx series. In Section 4, a more in depth look at the forecasting performances of HDBSTAR models is presented in the Brazilian electricity load series application. The article concludes with some discussion of this work in Section 5.

2 The harmonic DBSTAR model

A DBSTAR model in its simplest form can be seen as polynomial approximation of the classical STAR model as defined by (1), where a dynamic smooth transition function, $\pi(s_t; \gamma_t, c_t)$, similar to the logistic in (2) but with both the smoothing parameter γ and the threshold parameter c of $\pi(s_t; \gamma, c)$ allowed to change in time, is represented by a polynomial approximation. This paper considers two distinct approximations, the Taylor series expansion and the B-spline function, that characterise the Taylor and the B-spline DBSTAR models, respectively. Despite based on the logistic transition function, the development below can without loss be adopted for any other transition function that can be approximated by a polynomial function.

So, for a dynamic logistic transition function $\pi(s_t; \gamma_t, c_t)$ with real values in the interval $[0, 1]$, where s_t is a transition variable, $\gamma_t \in \Re^+$ is a smoothing parameter and $c_t \in \Re$ is a threshold value, a DBSTAR(r, p) model of orders r and p is defined by the set of quadruple $\{\underline{F}_t, \mathbf{G}_t, \Sigma_t, \mathbf{W}_t\}$ as follows. Note that underlined characters are used to represent vectors, boldface to represent matrices and prime to denote transposition throughout.

The observational and the system equations of a DBSTAR (r, p) are respectively given by

$$(Y_t \mid \underline{\theta}_t) \sim N\left(\underline{F}_t'\underline{\theta}_t, \Sigma_t\right) \tag{3}$$

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_t\underline{\theta}_{t-1}, \mathbf{W}_t\right) \tag{4}$$

where $\underline{F}_t' = [\underline{z}_t, B_1(s_t)\underline{z}_t, \ldots, B_{r-1}(s_t)\underline{z}_t, B_r(s_t)\underline{z}_t]$ is a *known* $(r+1)(p+1)$-dimensional vector of polynomial regression variables $B_i(s_t)\underline{z}_t$ $(i = 0, 1, \ldots, r)$ with $B_i(s_t)$ known functions of s_t which form depends on the approximation used for $\pi(s_t; \gamma_t, c_t)$ and $\underline{z}_t = (1, y_{t-1}, \ldots, y_{t-p})$; $\underline{\theta}_t$ is the state vector containing *unknown* parameters associated with the components of $\underline{F}_t$, i.e, $\underline{\theta}_t^{'} = (\underline{\theta}_0, \underline{\theta}_1, \ldots, \underline{\theta}_r)_t$ with elements $\underline{\theta}_{it} = (\theta_{i0}, \theta_{i1}, ..., \theta_{ip})_t$ where the elements of $\underline{\theta}_{0t}$ are $\theta_{0jt} = \phi_{1jt} + \beta_{0t}\phi_{2jt}$ $(j = 0, ..., p)$ and $\underline{\theta}_{it} = \beta_{it}\underline{\phi}_{2t}$ $(i = 1, ..., r)$ such that $\theta_{ijt} = \beta_{it}\phi_{2jt}$; and β_{it} $(i = 0, 1, ..., r)$ are polynomial functions of γ_t and c_t only such that

$$\pi(s_t; \gamma_t, c_t) \simeq \sum_{i=0}^{r} \beta_{it}(\gamma_t, c_t) B_i(s_t) \ . \tag{5}$$

Note that in this form, the parameters γ_t and c_t of $\pi(s_t; \gamma_t, c_t)$ are separated from the observable s_t such that they can be included together with the dynamic AR coefficients $\underline{\phi}_{it} = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{ip})_t$ where ϕ_{ijt} is the coefficient j $(j = 0, 1, \ldots, p)$ of the AR regime i $(i = 1, 2)$, into the *unknown* state vector $\underline{\theta}_t$ in (3) above, while s_t can be included together with past values of y_t into the *known* vector $\underline{F}_t$. For simplicity, the number of regimes is restricted to two AR(p) models here, albeit the methodology can be relatively straightforwardly extended to multiple regimes and differing AR orders.

The observational variance Σ_t in (3) is considered *unknown* and defined as $\Sigma_t = k_t V$, where $k_t = k(\mu_t)$ is an appropriately chosen variance law (a scaling function of the mean $\mu_t = \underline{F}_t'\underline{a}_t$ of Y_t, where $\underline{a}_t$ is the mean of the prior distribution of $\underline{\theta}_t$). V is the unknown variance scale parameter that is allowed to change stochastically. While a suitable chosen variance law can model systematic changes in the observational variability in time, we assume that Σ_t may change stochastically but only slowly and steadily in time (with the use of a variance discounting technique) to avoid potential unpredictable behaviour that can lead to loss of analytical tractability (Broemeling, 1985).

2.1 The Taylor DBSTAR model

A *Taylor DBSTAR* model is defined as the DBSTAR model for which $B_i(s_t)$ in (5) is a polynomial function of order i of the form

$$B_i(s_t) = s_t^i \tag{6}$$

and $\beta_{it}(\gamma_t, c_t)$ is obtained by expressing the Taylor series expansion of $\pi(s_t; \gamma_t, c_t)$ in the vicinities of $s_t = c_t$. So, a Taylor DBSTAR(r, p) model is characterised by the observational equation (3) above where the Taylor series expansion of the transition function is truncated at order r. Thus, at each time t a Taylor DBSTAR(r, p) model corresponds to a STAR model of order p which transition function is approximated by its Taylor expansion truncated at order r. Notice that the Taylor series expansion approximates $\pi(s_t; \gamma_t, c_t)$ better in the vicinities of $s_t = c_t$, so that, at each time t, the approximation changes for changes in c_t.

2.2 The B-spline DBSTAR model

A *B-spline DBSTAR* model on its turn is defined as the DBSTAR model for which the function in (5) is such that $B_i(s_t)$ is a B-spline basis function (a piecewise polynomial function) and $\beta_{it}(\gamma_t, c_t)$ are the associated coefficients. The B-spline of basis functions $B_i(s_t)$ of degree q, given a number n of knots in an interval is defined as in Wold (1974).

Computationally, the B-splines are obtained recursively by the Cox-de Boor algorithm (de Boor, 1978). As pointed out by Eilers and Marx (1996), they are rather attractive as base functions for univariate regression in which a linear combination of cubic B-splines gives a curve smooth enough to provide a good fit in many applications. This is the case here for the logistic transition function (2) (or for any of the usual alternative transition functions such the second-order logistic and the exponential). In our case, an appropriate order $r = q + n - 2$ in (5) is usually determined by the number n of knots chosen. In the splines regression context, the choice of n can be a complex task and statistics with penalties for overfitting are used in determining the optimal number. In our context, n (and q) is determined via model selection approach.

Some parameters in a Taylor or B-splines DBSTAR model are not treated as unknown parameters to be estimated but fixed a priori. The main reason for that is to preserve analytical tractability in the Bayesian parametric updating that allows fast sequential computations. For the Canadian lynx trappings and electricity loading applications, optimal values of those parameters were obtained via model selection approach.

2.3 The HDBSTAR model

Now, in order to allow the modelling of any observed cyclic behaviour in terms of cyclical components explicitly, we introduce the *Harmonic* DBSTAR (HDBSTAR) model. Similarly to seasonality, the explicit modelling of long term cyclic behaviour allows accounting for changes in that behaviour in a forecasting model. Fourier form representations of the periodic behaviour that allows for modelling changes in amplitude and phase for fixed number of harmonics preserving, thus, the analytical tractability of parametric posterior distributions and forecasting functions are adopted here.

HDBSTAR models extend the set of quadruple $\{\underline{F}_t, \mathbf{G}_t, \Sigma_t, \mathbf{W}_t\}$, with $\underline{F}_t = (\underline{F}_{1t}, \underline{F}_{2t})$, $\mathbf{G}_t = (\mathbf{G}_{1t}, \mathbf{G}_{2t})$ and $\mathbf{W}_t = (\mathbf{W}_{1t}, \mathbf{W}_{2t})$, where $\underline{F}_{1t}, \mathbf{G}_{1t}$ and $\mathbf{W}_{1t}$ are associated to the nonlinear autoregressive components as in (3) and (4), and $\underline{F}_{2t}, \mathbf{G}_{2t}$ and $\mathbf{W}_{2t}$ are associated to the cyclical component. A HDBSTAR(r, p, h) model for cycles is defined as a DBSTAR(r, p) with an explicit component for cycle with h harmonics as follows

$$\left(Y_t \mid \underline{\theta}_t, \underline{\psi}_t\right) \sim N\left(\underline{F}_{1t}^{'}\underline{\theta}_t + \underline{F}_{2t}^{'}\underline{\psi}_t, \Sigma_t\right) \tag{7}$$

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{1t}\underline{\theta}_{t-1}, \mathbf{W}_{1t}\right) \tag{8}$$

$$\left(\underline{\psi}_t \mid \underline{\psi}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{2t}\underline{\psi}_{t-1}, \mathbf{W}_{2t}\right) \tag{9}$$

where $\underline{\psi}'_t = [\underline{\psi}_{1t}, ..., \underline{\psi}_{ht}]$, $\underline{\psi}_{jt} = (a_j, b_j)_t$ with a_{jt} and b_{jt} being the unknown Fourier coefficients of each harmonic $S_j(t) = a_{jt}cos(\omega_j t) + b_{jt}sin(\omega_j t)$ $(j = 1, ..., h)$. The $2h$-dimensional vector $\underline{F}_{2t}$ is a canonical partitioned vector associated to the harmonics in $\underline{\psi}_t$, with 1 in an harmonic position and 0 otherwise. For example, $\underline{F}_{2t} = [1, 0]$ for $h = 1$ harmonic, $\underline{F}_{2t} = [1, 0, 1, 0]$ for $h = 2$ harmonics, and so forth. The frequency of each harmonic is $\omega_j = 2j\pi/\tau_c$, where τ_c is the period of the cycle. The evolution matrix $\mathbf{G}_{2t}$ of the cyclical component is a block diagonal matrix $\mathbf{G}_{2t} = diag(\mathbf{H}_{1t}, ..., \mathbf{H}_{ht})$ where $\mathbf{H}_{jt}$ is the harmonic matrix with trigonometric elements such that $|\mathbf{H}_{jt}| = sin^2(w_j t) + cos^2(w_j t) = 1$. The $(2h \times 2h)$-matrix $\mathbf{W}_{2,t}$ contains the covariances of the cyclical components.

The first harmonic, the fundamental harmonic, is expected to dominate the cyclical pattern, having a strong sinusoidal signal. The higher frequency harmonics oscillate faster than the fundamental one and more appropriate for modelling higher frequency repetitive behaviour. Obviously that the larger the h the more accurate the modelling of periodic variations in the data. However, adopting the parsimony principle we look for the smallest h that can still provide a good representation of the cyclical component of the underlying process. For cases where a large enough initial dataset is available (as are the cases in this paper) to enable the investigation of the cyclic behaviour, an optimal value of h can be determined, for example, by a stepwise model selection approach.

Two types of DBSTAR models are defined here, the Taylor DBSTAR models based on approximating the adopted transition function by a Taylor series expansion, an the B-splines DBSTAR models which use B-spline functions for that purpose. B-splines are constructed from piecewise polynomial functions joined at knots created by dividing the underlying interval into parts. Those functions satisfy weak differentiability conditions that guarantee the continuity and smoothness of the resulting function. The reader can refer to de Boor (1978) and Eilers and Marx (1996) amongst others for more details on B-splines.

3 Main results of the Canadian lynx application

The Canadian lynx dataset is a yearly series with 114 observations of the number of lynx trapped in the Mackenzie River, district of North-west Canada, from 1821 to 1934. They were collected to improve knowledge about the population dynamics of the ecological system in that area. They have been used in various studies (see e.g. Terasvirta, 1994; Lopes and Salazar, 2005) to analyse and compare the fitting of proposed models. The most famous features of that series are (a) the lack of trend, (b) the presence of irregular changes in the amplitude in time, and (c) the presence of persistent non-regular cyclic oscillations with periods of 10 or 11 years. Those features have been familiar to biologists for a long time and are well known in historical records of trappings of lynx in Canada as described by Elton and Nicholson (1942).

Similarly to other studies of this series, including Chan and Tong (1986) and Lopes and Salazar (2005), the original series was $\log_{10}$-transformed here to remove the marked right-skewness of the data as well as to allow the comparative analysis with the classical STAR and the BSTAR models. The transformed series presented no evident trend but a clear periodic repetitive behaviour with significant estimated autocorrelations at lags

1, 2 and 11 in the partial auto-correlation function (PACF). However, it may not be appropriate to use the PACF for model order identification in this case as a graphical analysis of the scatterplots of y_{t-u} against y_{t-v} (for $u, v = 1, 2, ..., v > u$.) showed lack of linearity for most lags of the series. Consistently with Elton and Nicholson (1942), a periodogram of the series displayed a spike around the 0.1 frequency indicating cyclical behaviour with a 10-year wavelength.

Following a numerical grid search for optimal values (based on the log-smoothing likelihoods of the models calculated conditionally on values of d, p, r and h) the following three models were selected: a Taylor DBSTAR$(3, 12)$, a cubic B-splines DBSTAR$(6, 12)$ with five knots and a cubic B-splines HDBSTAR$(6, 2, 2)$ also with five knots. Similarly to Terasvirta (1994) and Lopes and Salazar (2005), the optimum delay parameter of the adopted dynamic logistic transition function was found to be $d = 3$ for all models, that is $s_t = y_{t-3}$. However, unlike both Lopes and Salazar (2005) and Terasvirta (1994) that found $p = 11$, the optimum AR order for the non-harmonic Taylor and B-splines DBSTAR models were found to be $p = 12$. The optimal harmonic model though was found to be considerably more parsimonious. In fact, the optimal values for the number of harmonics and AR order were $h = 2$ and $p = 2$ respectively for the B-splines HDBSTAR model. A Taylor expansion of order $r = 3$ was adopted for the Taylor DBSTAR$(3, 12)$ model as higher orders only provided marginal improvements.

Table 1 displays the MAE and the RMSE of the fitted logistic classical STAR(11) from Terasvirta (1994) and the logistic BSTAR(11) from Lopes and Salazar (2005) as well as those of the selected Taylor DBSTAR$(3, 12)$, the B-splines DBSTAR(6, 12) and the B-splines HDBSTAR$(6, 2, 2)$ models. It is clear that the B-splines HDBSTAR(6, 2, 2) model with a MAE of 0.006 and a RMSE of 0.013 produced the best fit of all models. Its MAE is 20% lower than the second best fit of the Taylor DBSTAR(3, 12) although its RMSE is only marginally lower. The B-splines DBSTAR(3, 12), with the largest MAE and RMSE of all DBSTAR models, still outperformed the classical STAR(11) and the BSTAR(11) models, by quite a margin (its MAE and RMSE were 9.8 and 5.7 times lower than the BSTAR(11) of Lopes and Salazar (2005), respectively). A static version of the Taylor DBSTAR(3,12) model (with discount factors set to unity) had an MAE and an RMSE of 0.109 and 0.141, respectively, produced a fitting that was only marginally better than the BSTAR(11)

Model	MAE	RMSE
Lopes and Salazar (2005) - BSTAR(11)	0.118	0.153
Terasvirta (1994) - STAR(11)	0.142	0.179
Taylor DBSTAR$(3, 12)$	0.012	0.015
B-splines DBSTAR$(6, 12)$	0.014	0.027
B-splines HDBSTAR$(6, 2, 2)$	**0.006**	**0.013**

Table 1: Mean Absolute Errors (MAE) and Root Mean Squared Errors (RMSE) of compared models

As seen above, the sequential dynamic modelling of the Canadian lynx data by the DBSTAR models have allowed much improved fitting to the data compared with the static STAR and BSTAR models. This sequential modelling also allows a better understanding of the lynx population dynamics via analysis of the obtained dynamic parameters estimated from the data that helps to verify how various components varied in time. For example, the estimated cyclic component (obtained from the posterior means of the harmonic components of the HDBSTAR model by Kalman fitering) has shown increased variability in amplitudes from 1824 to 1846, followed by lower oscillations until approximately 1896 after when oscillations increased again almost mirroring the initial period.

It is worth mentioning that a residual analysis of the DBSTAR models above showed residuals to be uncorrelated as well as no significant departures from normality. Please refer to Santos (2014) and Faria and Santos (2018) for more details of the application above.

4 Electricity load forecasting in southern Brazil

The hourly series of electricity load, measured in MegaWatts (MW), and temperature, in degree Celsius (°C), are from the Southeast and Central-West regions of Brazil and span from the first hour on 1 June 2003 to the last hour on 30 June 2010 (that is, 62,088 hourly observations covering the 7-year period). The hourly load data are aggregated for the region while the hourly temperature data are averaged across all the states in those regions. Calendar variables indicating weekdays, national holidays and bridge-holidays (i.e. days between midweek bank holidays that are near a weekend) were used in the models to account for their effects on the load. To measure the short-term forecasting performances of the selected DBSTAR models, the initial 61,368 observations (covering the period from 1 June 2003 to 31 May 2010) and the last 720 observations (corresponding to one month of hourly data from 1 June 2010 to 30 June 2010) were used as in-sample and out-of-sample data respectively.

A preliminary analysis of the data has shown, amongst others, a non-linear S-shaped relationship between load and temperature. In this relationship, differently from the U-shape normally observed for northern hemisphere countries, the load tends to increase with increases in temperature although on a non-linear fashion resembling an S-shape. In fact, the rate of increase in weekly average load is lower at lower levels of weekly average temperature (16-20°C) as compared with higher levels (26-30°C) when higher rates of increase occur, suggesting that at least two regimes of distinct consumption behaviour is present in the load series that can justify at least in part the use of STAR type of models. Other marked characteristics observed from the data analysis are the long-term positive trend with slowly increasing variability in time, and both within-day and within-week changing periodic behaviour and variability. The within-day variation of load has a general shape that is similar for all weekdays with lower loads in the early hours with minimum levels between 4-8am followed by sharp increases until the middle of the day when they stabilise at an intermediate level until early afternoon. This is followed by peaks in the evening between 7-9pm. Sundays and bridge-holidays show a little shift of consumption to the right in the hour scale as well as lower levels at almost all times.

Similarly, Saturdays and holidays tend to group together with a smaller shift to the right but with larger loads in the early hours. Except for Monday (with slightly lower loads at all times except in the early hours when they are lowest), the patterns for other weekdays are practically undistinguishable from each other. They show the largest levels of load overall with troughs typically at 4-5am, lower peaks at 11am-12pm and 3-4pm and the largest peaks from 7-9pm. The within-week variations are such that daily peaks change their patterns (or shape) along the year according to the season. In particular, evening peaks observed in the winter is not present in the during summer when observed smoother peaks are thought to be influenced by the availability of natural lights early in the evenings, helped by the change in the Brazilian summer time (when clocks are moved forward by one hour from October to February). There is also a strong within-year seasonal effect on load that is very much in line with the strong seasonal behavior of the temperature. It is also worth pointing out that due to the changing variability, both within-the-day (that is larger at peak than at off-peak hours) and within-the-week (that is larger at working days than at weekends and holidays), the underlying load series presented intrinsic second-order non-stationarity that no differences and/or transformation in the class of Box-Cox transformations could be found to turn the series stationary. No serious departures from normality were observed.

The Taylor and the B-splines HDBSTAR models that were formulated in this application to account for the above described characteristics of the electricity load series include a trend and a seasonal component, two cyclical components (one for the within-day and one for the within-week periodic variations), one calendar component and one non-linear AR component with temperature used as the transition variable (associated with a logistic STAR) to account for the non-linear effects of temperature on load. Note that the cyclical components aim to account for the non-linear cyclic behaviour not accounted for by the seasonal component.

A Taylor HDBSTAR(3, 1, 2) model and a cubic B-splines HDBSTAR(3, 1, 2) model with $n = 1$ knot (located at the median temperature of 23°C) that included the components mentioned above, with delay $d = 1$ for the temperature series as transition variable, were selected as the optimal models. They presented the largest in-sample conditional log-predictive likelihoods in a grid search of models of varying orders. Initial non-informative prior distributions were attributed to the hyper-parameters of the models. Discount factors of $\delta_V = 0.90$ and $\delta_W = 0.99$ were also found as optimal values. Two harmonics for each cyclical component were found to represent the non-linear short-term periodic patterns adequately.

The B-splines HDBSTAR(3, 1, 2) model fitted the data slightly better during the in-sample period than the Taylor HDBSTAR(3, 1, 2) model with a larger log-posterior likelihood (of -559751 against -550437), lower MAPE (0.00527 against 0.00547) and RMSE (208.19 against 227.68).

Both the Taylor HDBSTAR(3, 1, 2) and the cubic B-splines HSDBSTAR(3, 1, 2) models were also used to produce rolling forecasts sequentially from 1 to 72-hours-ahead horizons during the out-of-sample period. For that, forecasts of temperatures from a multiplicative Winters model (selected by SPSS's expert model amongst a number of exponential smoothing and ARIMA models) were used. Those DBSTAR models were

implemented in R and used prior-to-posterior updating and forecasting routines based on the Kalman filter (and on the DLM and SPLINES packages). Their running times, for fitting and forecasting on a desktop PC with i7 processor at 3.30GHz with 32GB of memory and SSD disk drive, varied from 8.93 (for a forecasting horizon $h = 24$ hours-ahead) to 14.50 minutes (for $h = 72$) for the Taylor HDBSTAR model and from 10.23 ($h = 24$) to 17.65 minutes ($h = 72$) for the B-splines HDBSTAR model.

Table 2 shows the MAPE (%) and the RMSE for the Taylor and the B-splines models for each forecasting horizon of 1, 12, 24, 48 and 72 hours-ahead. The values in bold represent the lowest values between the two models for each horizon. Note that the B-splines model outperformed the Taylor model for 1 and 12-hours-ahead horizons but for the 24, 48 and 72-hours-ahead horizons the Taylor outperformed the cubic B-splines model on both MAPE and RMSE criteria. Overall, with only a few exceptions, the differences in forecasting performances were relatively small with both models displaying fairly similar performances.

For illustration, the MAPE and RMSE for a SARIMA(2,1,10)(2,1,1) fitted by SPSS as the best classical forecasting model using the temperature series and the calendar indicators as explanatory variables were 107.80 % and 455.993 respectively. As expected, those results compare unfavourably with the much smaller MAPE and RMSE of 59.06 % and 221.29, respectively, for the B-splines HSDBSTAR(3,1,2) and 61.20 % and 231.77 for the Taylor HSDBSTAR(3,1,2), indicating improved fitting of nearly 50% by the DBSTAR models.

Horizon	Model	In-sample		Out-of-sample	
		MAPE (%)	RMSE	MAPE (%)	RMSE
1	Taylor	61.20	231.77	54.65	227.68
	B-splines	**59.06**	**221.29**	**52.74**	**208.19**
12	Taylor	210.28	915.90	173.62	804.68
	B-splines	**111.90**	**401.20**	**119.06**	**448.55**
24	Taylor	204.27	**890.51**	**184.98**	**850.83**
	B-splines	**204.22**	893.91	199.29	877.58
48	Taylor	**203.29**	**889.03**	**186.95**	**876.09**
	B-splines	204.18	894.67	201.75	897.90
72	Taylor	**197.07**	**880.96**	**193.12**	**897.88**
	B-splines	198.72	887.22	209.41	925.56

Table 2: Mean Absolute Percentage Errors (MAPE) and Root Mean Squared Errors (RMSE) of the Taylor and the B-splines HDBSTAR(3, 1, 2) models for each of the 1, 12, 24, 48 and 72 hours-ahead forecasting horizons

For the 24-hours-ahead forecasting horizon, that is usually of particular interest to practitioners, plots of observed loads versus forecasts during the out-of-sample period,

displayed most points close to a diagonal line indicating the generally good forecasting accuracy for both the Taylor and B-splines HDBSTAR(3,1,2) models. However, the B-splines model showed a slightly higher degree of scatter than the Taylor model as expected from their MAPE and RMSE out-of-sample performance measures for the 24-hours horizon in Table 2.

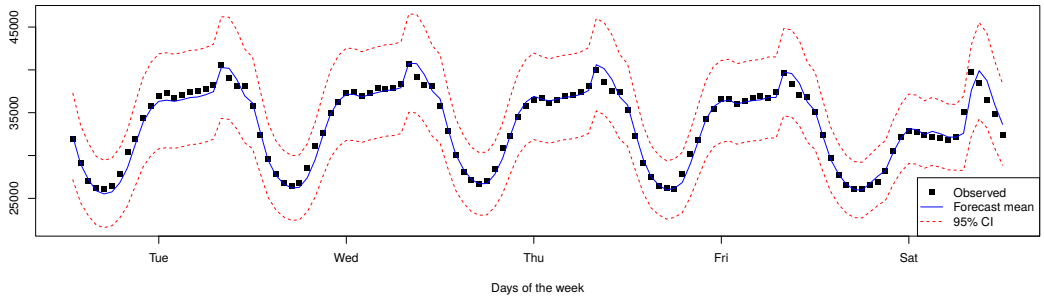
Figure 1 (a)–(d) shows the observed loads (solid square points), their 24-hours-ahead rolling forecasts (solid line) and 95% predictive intervals (dashed lines) during the out-of-sample period by the Taylor HDBSTAR(3, 1, 2) model. At each hour during the out-of-sample period, the forecast at that hour was made 24-hours earlier using the 24-hour-ahead temperature forecast as the transition variable. Those temperature forecasts were obtained by a Winters' exponential smoothing models with parameters $\alpha = 0.973$ for level and $\delta = 0.07$ for seasonal (the trend parameter $\gamma = 0.001$ was non-significant with a p-value of 0.387).

Note that, in general, the forecasts are quite close to the observed loads with 95% prediction intervals that are particularly larger at peak times (18-21 hours) than at late night and early morning times in most days. The forecasts were well within the bounds of the prediction intervals at almost all times. The selected out-of-sample period of June 2010 had a number of special days such as the Corpus Christie holidays on Thursday and Friday, 3^{rd} and 4^{th} June, in Figure 1(a), and the South African football world cup when on Tuesday, 15^{th} June, in Figure 1(b), Brazil played North Korea at 15:30 hours as well as at 09:00 on Sunday, 20th June, and at 11:00 on Friday, 25^{th} June, both in Figure 1(c) when Brazil played Ivory Coast and Portugal respectively, followed by Chile at 15:30 on Monday, 28^{th} June, in Figure 1(d). It can noticed that on those events the electricity load decreased comparatively with similar times at similar days. This can be explained by the fact that large amounts of people in urban areas tend to group together with family members on religious holidays and with friends in bars and restaurants or with crowds of people in public places such as squares with large screens to watch the national football team. Notice that the effect of the football match events lasted mainly on the hours of those events with the load levels increasing back to higher levels soon after.

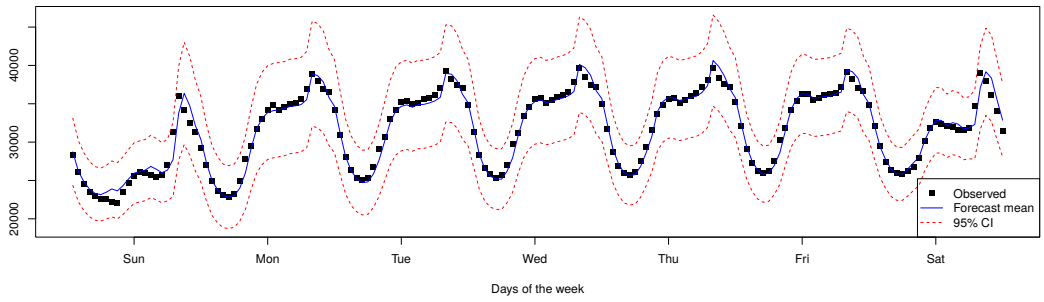
In practice, an expert user could anticipate effects of events like those above in similar occasions and make appropriate interventions in the model (by changing prior hyper-parameters accordingly).

5 Concluding remarks

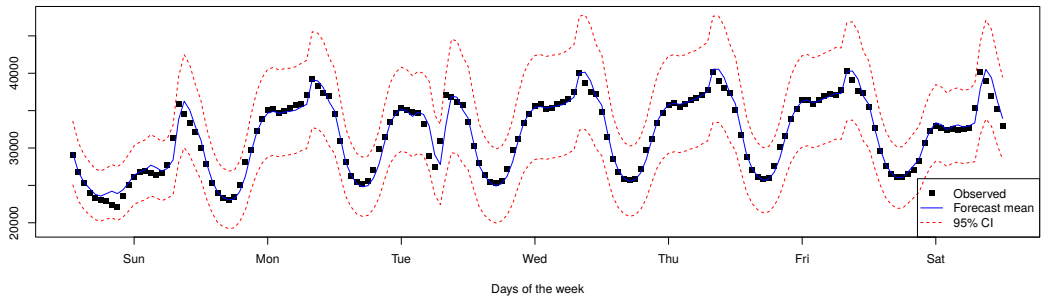
The *dynamic Bayesian smooth transition autoregressive* (DBSTAR) models introduced here are Gaussian analytical approximations of STAR models based on polynomial *dynamic linear models* (DLMs) of West and Harrison (1997). They are appropriate for non-linear and intrinsically non-stationary auto-regressive time series processes such as those exhibiting asymmetric cycles and offer an alternative to both the classical STAR models of Chan and Tong (1986) and the *computational Bayesian STAR* (BSTAR) models of Lopes and Salazar (2005). Two types of DBSTAR models, the Taylor and B-splines DBSTAR models, were defined according to the adopted approximation approach adopted for functional logistic transition functions associated with STAR models. Their unknown



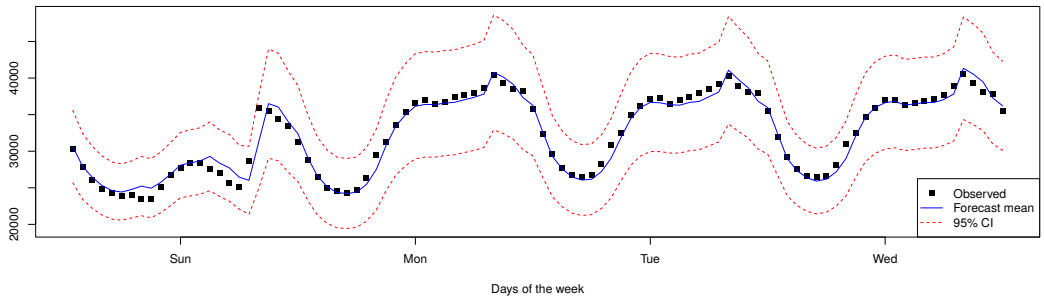
(a) 1-5 June 2010



(b) 13-19 June 2010



(c) 20-26 June 2010



(d) 27-30 June 2010

Figure 1: Out-of-sample observed values (dots), the 24-hours-ahead forecasts (solid line) and the corresponding 95% prediction intervals (dashed lines) of the hourly load by the Taylor HDBSTAR(2, 1, 2) model from the first to the last hour in the periods (a) 1-5 June 2010; (b) 13-19 June 2010; (c) 20-26 June 2010; and (d) 27-30 June 2010.

parameters are sequentially updated in time analytically via Kalman filtering approach as described in Faria and Santos (2018). Similar to classical STAR models some parameters such as the AR order and transfer function location values are assumed fixed a priori. However, unlike both the classical STAR and the BSTAR the unknown parameters (AR coefficients, observational variance and transition smoothing variable) are all dynamic. In applications where initial data are available optimisation approaches can be used to determine suitable values for the fixed parameters. The DBSTAR model formulation is thus useful in applications demanding sequential parametric change in time (including observational variance) and fast computing. However, a complicating factor in the DBSTAR model formulation is that the solution of a set of polynomial equations is required at each time step if the first two moments of the unknown parameters associated with a STAR model are to be determined. This is because, at each time step, a DBSTAR model produces posterior distributions for the parameters of the polynomial DLM that results from approximating the STAR transition function by Taylor expansion or B-splines. Those parameters are polynomial functions of the original STAR parameters. Thus, a DBSTAR model naturally estimate parameters that have polynomial auto-regression interpretability of their own but will demand extra computational processing for parametric interpretability associated with parameters of a STAR model.

Similarly to DLMs, DBSTAR models can be formulated to account for components of the underlying process. Trend, seasonality and cycle components are easily accounted for in a parsimonious way. Heteroskedasticity can also be accounted for by either incorporating a variance law in the model or with the use of variance discounting techniques when slow but steady changes in the unknown observational variance are allowed. The parameters associated with with the slow changes are estimated sequentially from data. To model observed cyclical behaviour in the data that are not accounted for by seasonal components, *harmonic* DBSTAR (HDBSTAR) models have been defined that explicitly include components for cycles. Fourier form representations of cycles with combinations of sine/cosine waves provide an economic parametric characterisation and facilitate their interpretation. In general, lower auto-regressive orders are required by a HDBSTAR model comparatively to a DBSTAR (and a STAR) model. The parsimony of a HDBSTAR model is balanced by larger amplitudes in the autoregressive coefficients. This is an advantage of the HDBSTAR models over the DBSTAR, classical STAR and BSTAR, for modelling time series in the presence of repetitive periodic behaviour.

Taylor and the B-splines formulations of HDBSTAR models were applied to a large hourly series of electricity load in a region in Brazil. In this application, the formulated HDBSTAR models showed considerably improved fitting over a SARIMA model obtained by SPSS's expert forecasting modeler. The B-splines slightly outperformed the Taylor HDBSTAR model for both fitting and short-term forecasting. The Taylor model performed slightly better for longer forecasting horizons.

References

Bauwens, L., Lubrano, M., and Richard, J. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.

Broemeling, L. D. (1985). *Bayesian analysis of linear models*. Decker, New York.

Campbell, E. P. (2004). Bayesian selection of threshold autoregressive models. *Journal of Time Series Analysis*, 25:467–482.

Chan, K. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7:179–190.

Chen, C. and Lee, J. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, 16:483–492.

Chen, C. W. (1998). A bayesian analysis of generalized threshold autoregressive models. *Statistics & Probability Letters*, 40:15–22.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer, Berlin.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties (with discussions). *Statistical Science*, 11:89–121.

Elton, C. and Nicholson, M. (1942). The ten-year cycle in numbers of the lynx in canada. *Journal of Animal Ecology*, 11:215–244.

Faria, A. and Santos, A. (2018). Dynamic bayesian smooth transition auto-regressive models for non-linear non-stationary time series. *The Open University, Statistics Technical Report*, 18/04:1–28.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

Geweke, J. and Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, 14:441–454.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732.

Lopes, H. F. and Salazar, E. (2005). Bayesian model uncertainty in smooth transition autorregressions. *Journal of Time Series Analysis*, 27:99–117.

Lubrano, M. (2000). Bayesian analysis of nonlinear time-series models with a threshold. *Nonlinear Econometric Modelling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory*.

Santos, A. (2014). *Dynamic Bayesian Smooth Transition Autoregressive models for non-stationary nonlinear time series*. The Open University, PhD Thesis.

Terasvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218.

Tong, H. (1978). *On a threshold model*. In Patern Recognition and Signal Processing.

Tong, H. (2011). Threshold models in time series analysis - 30 years on (with discussions). *Statistics and Its Interface*, 4:107–136.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer.

Wold, S. (1974). Spline functions in data analysis. *Technometrics*, 16:1–11.

CP-based cloud workload annotation as a preprocessing for anomaly detection using deep neural networks

Gilles Madi Wamba, and Nicolas Beldiceanu

Institut Mines-Telecom Atlantique, LS2N, Nantes, France
last-name.first-name@imt-atlantique.fr

Abstract. Over the last years, supervised learning has been a subject of great interest. However, in presence of unlabelled data, we face the problem of deep unsupervised learning. To overcome this issue in the context of anomaly detection in a cloud workload, we propose a method that relies on constraint programming (CP). After defining the notion of quasi-periodic extreme pattern in a time series, we propose an algorithm to acquire a CP model that is further used to annotate the cloud workload dataset. We finally propose a neural network model that learns from the annotated data to predict anomalies in a cloud workload. The relevance of the proposed method is shown by running simulations on real-world data traces and by comparing the accuracy of the predictions with those of a state of the art unsupervised learning algorithm.

1 Introduction

With the increase of the use of cloud computing technologies, more and more applications rely on a cloud infrastructure. This model is named Infrastructure as a Service (IaaS). Since failures are common in such large systems [21], the cloud provider needs to ensure the integrity of the infrastructure, prevent failures or fix them as soon as possible.

Most predictive machine learning algorithms rely on historical data to learn a predictive model that can predict the outcome of a given observation. Generally, cloud providers have large quantity of historical data obtained by monitoring some metrics. Such cloud workloads are modelled by time series that are generally not labelled. To address this issue, some data scientists require human intelligence through crowdsourcing services like Amazon Mechanical Turk [20, 17]. The drawback of such approach is its cost in time and money, and even more the reliability of the produced annotations [13]. In the absence of labels, the problem of deep unsupervised learning has been studied in [10]. Most of these approaches rely on k-means clustering [14]. As shown in Section 4.2, a clustering approach is not well suited for anomaly detection in the context of cloud workloads.

In this context, the contribution of this paper is twofold. First we introduce the generic notions of *extreme patterns* (i.e. pattern occurrences within a time

series for which one or several feature values take very small or very large values relatively to the rest of the time series), and of *quasi-periodic pattern group* within a time series. Second, we propose a three-step method to solve the problem of predicting malfunction in a cloud data centre infrastructure given its workload:

- Using the notion of *extreme patterns*, we first acquire a constraint programming (CP) model that is able to discriminate between a workload presenting a malfunction and one that does not.
- Then we use the acquired CP model to automatically label the historical workload dataset.
- Finally, we train a deep neural network to predict the malfunctions.

The rest of the paper is organised as follows. Section 2 provides the necessary background on time series in the context of constraint programming. Section 3 details our main contribution, namely the definition of *quasi-periodic extreme patterns* and the illustration how to use it to annotate cloud workloads. In Section 4, we present two methods, the first one based on supervised learning, and the second one based on unsupervised learning (through the k-means algorithm) to predict anomalies in a cloud workload. Finally, we evaluate and compare both models in Section 5 and conclude in Section 6.

2 Background

A *time series* is a sequence corresponding to measurements taken over time. We use time series to model resource production or consomption, e.g. energy, memory, cpu, bandwidth [6, 15]. These type of time series are constrained by physical or organizational (in general structural) limits as well as by business rules, which restrict the evolution of the series.

Definition 1 (signature). *Given a time series* $x = x_1, x_2, \ldots, x_n$*, the* signature *of* x *is the sequence* $s = s_1, s_2, \ldots, s_{n-1}$*, where every* s_i *is defined by* $(x_i < x_{i+1} \Leftrightarrow s_i = \text{`}<\text{'}) \wedge (x_i = x_{i+1} \Leftrightarrow s_i = \text{`}=\text{'}) \wedge (x_i > x_{i+1} \Leftrightarrow s_i = \text{`}<\text{'})$.

Definition 2 (s-occurrence, i-occurrence, e-occurrence, found index). *Consider a time series* $x = x_1, x_2, \ldots, x_n$*, its signature sequence* $s = s_1, s_2, \ldots, s_{n-1}$*, a regular expression* σ *over the alphabet* $\{\text{`}<\text{'}, \text{`}=\text{'}, \text{`}>\text{'}\}$*, two natural numbers* b *and* a*, and a subsignature* $s_i, s_{i+1}, \ldots s_j$ *with* $1 \leq i \leq j \leq n-2$*, forming a maximal word in* s *matching* σ*. The* s-occurrence *of* σ *is the index sequence* $i, \ldots, j$*; The* i-occurrence *of* σ *is the index sequence* $i + b, \ldots, j$*; The* e-occurrence *of* σ *is the index sequence* $i + b, \ldots, j + 1 - a$*; The* found index *is the smallest index* k *in the interval* $[i, j]$ *such that the word* $s_i s_{i+1} \ldots s_k$ *is an occurrence of* σ*.*

An s-occurrence identifies a maximal occurrence of a pattern in a signature sequence, while an i-occurrence identifies a maximal occurrence of a pattern in an input sequence. Note that, by Definition 2, i-occurrences of the same pattern never overlap. The feature value of a pattern occurrence is computed from the e-occurrence, where the constants b and a are used for respectively trimming the left and the right borders of the regular expression σ.

3 Detection of quasi-periodic extreme patterns in CP

Housssam *et al.* present in [2] a formalisation of arrhythmia detection algorithms. Their algorithm relies on quantitative regular expressions [3] for peak detection. Scholkmann *et al.* [18] propose another parameter free algorithm for peak detection in periodic or quasi-periodic signals. While those algorithms are designed for peak detection only, our framework [5, 11] is compatible with a large class of regular expressions [11] and offers more than 700 time-series constraints [4].

First, since we aim at extracting relevant maximal occurrence of patterns, we formally define the notion of extreme pattern occurrences. Second, since we also aim at quantifying how regularly such maximal occurrence of patterns appear we define the notion of quasi-periodic extreme pattern. All the corresponding definitions will be illustrated on a concrete example at the end of this section.

3.1 Defining extreme pattern occurrences

In the same way we already synthesise a register automaton for computing the result associated with a time-series function, we also synthesise a register automaton for computing the feature values of each maximal occurrence of a given pattern in a time series in linear time wrt the time-series length. This will be used to extract the extreme feature value, a notion that we now define.

Definition 3 (extreme feature value of a pattern occurrence). *Given a regular expression σ, a feature f, a time series ts containing at least one maximal occurrence of σ, and a threshold $\tau \in [0, 1]$, the i^{th} maximal occurrence of σ within ts is* extreme *iff*

$$f_\sigma(i) \in [\min_{f,\sigma}(ts), \quad \min_{f,\sigma}(ts) + \tau \cdot \min_{f,\sigma}(ts)] \vee$$
$$f_\sigma(i) \in [\max_{f,\sigma}(ts) - \tau \cdot \max_{f,\sigma}(ts), \ \max_{f,\sigma}(ts)], \text{ where}$$

- $f_\sigma(i)$ *is the feature value of i^{th} maximal occurrence of σ,*
- $\min_{f,\sigma}(ts)$ *(resp.* $\max_{f,\sigma}(ts)$*) is the minimum (resp. maximum) feature value among all maximal occurrences of σ within ts.*

Definition 4 (extreme pattern occurrence). *Given a regular expression σ and a set of distinct features $f_1, f_2, \ldots f_k$, a maximal occurrence of σ in a time series ts is an* extreme occurrence *of σ wrt $f_1, f_2, \ldots f_k$ iff all corresponding feature values are extreme feature values for σ in ts.*

3.2 Identifying quasi-periodic extreme patterns

Definition 5 (footprint). *Given a time series $ts = x_0, x_1, \ldots, x_{n-1}$ and a pattern p, the* footprint *of p over ts is the sequence $fp = q_0, q_1, \ldots, q_{n-1}$ of the same length as ts where each $q_k \in fp$ is an integer between 0 and n such that:*

- $q_k = 0$ *if index k does not occur in any i-occurrence of the pattern p in the input time series ts.*

– $q_k = j > 0$ *if index* k *belongs to the* j^{th} i*-occurrence of the pattern* p *when reading the time series* ts*.*

Definition 6 (footprint distance). *The* footprint distance *between the* j^{th} *and the* $(j+1)^{th}$ *maximal occurrence of a pattern* p *in a time series* ts *is the number of zeros in the footprint of* p *over* ts *between the end of the* j^{th} i*-occurrence of* p *in* ts *and the start of the* $(j+1)^{th}$ i*-occurrence of* p *in* ts*.*

Definition 7 (found distance). *The* found distance *between the* j^{th} *and the* $(j+1)^{th}$ *occurrence of a pattern* p *in a time series* ts *is the difference between the two corresponding found indices.*

Definition 8 (distance). *Consider a pattern* p *for which the length of its smallest occurrence is* $\ell > 0$*. The* distance *between the* j^{th} *and the* $(j+1)^{th}$ *maximal occurrence of a pattern* p *in a time series* ts *is the footprint distance iff, given any occurrence* o *of the pattern* p*, all prefixes and suffixes of* o *for which the minimum length is greater than or equal to* ℓ *are also occurrences of the pattern* p*, and the found distance otherwise.*

Definition 9 (quasi-periodic pattern of order g). *Let* g *and* δ *be two natural numbers with* $g > 0$*,* p *a pattern,* ts *a time series and* o_p *the sequence of all maximal occurrences of* p *in* ts*. In addition let* e_p *be a subsequence of* o_p *of size at least* $\ell = 3g$*. Let* $\underline{d_p}$ *and* $\overline{d_p}$ *be resp. the minimum and the maximum distance between any* $j^{\texttt{th}}$ *and* $(j+g)^{\texttt{th}}$ *maximal occurrences of* p *in* e_p*, with* $i \in [1, \ell - g]$*. Then* p *is said to be a* quasi-periodic pattern *of order* g *and period* $[\underline{d_p}, \overline{d_p}]$ *iff*

1. $\overline{d_p} - \underline{d_p} \leq \delta$,
2. *the first maximal occurrence of* $p \in e_p$ *is located at a distance* $d \leq \overline{d_p}$ *from the beginning of* ts*,*
3. *the last maximal occurrence of* $p \in e_p$ *is located at a distance* $d' \leq \overline{d_p}$ *from the end of* ts*.*

When e_p corresponds to all maximal occurrences (resp. all extreme occurrences) of p then p is a periodic pattern (resp. periodic extreme pattern) of order g.

Example 1. We illustrate the previous definitions on the time series $0, 2, 9, 8, 10, 1, 0, 0, 3, 0, 8, 1, 9, 1, 1, 1, 2, 9, 9, 2, 2, 8, 0, 1, 0$. For this purpose we consider the PEAK (resp. STEADY_SEQUENCE) pattern defined by ' $< (= \mid <)^* (> \mid =)^* >$ ' (resp. ' $=^+$ ') and by the trimming parameters $b = a = 1$ (resp. $b = a = 0$). The eight (resp. four) subsequences outlined in yellow between two vertical red bars on top (resp. at the bottom) of Figure 1 correspond to the eight s-occurrences of the PEAK (resp. STEADY_SEQUENCE) pattern.

The distance of the PEAK pattern is defined as the found distance since not all large enough suffixes of a maximal occurrence of peak correspond to a peak, i.e. the suffix ' $>>$ ' of the second maximal occurrence of peak ' $<>>$ ' is not a peak itself. Consequently the distances $2, 4, 2, 2, 6, 3, 2$ between consecutive maximal occurrences of peaks are defined as the difference between their corresponding found positions, outlined with a vertical arrow in the upper part of Figure 1.

The distance of the STEADY_SEQUENCE pattern is defined at the footprint distance since all large enough prefixes and suffixes of ' $=^+$ ' also correspond to a steady sequence. Consequently the distances $6, 2, 1$ between consecutive maximal occurrences of steady sequences are defined as the difference between the start of a maximal occurrence of the pattern and the end of the maximal occurrence of the previous pattern.

We now focus on the extreme occurrences of the PEAK pattern wrt the max feature assuming a threshold $\tau = 0.2$. Such feature values $9, 10, 3, 8, 9, 9, 8, 1$ are shown in the lower part of Figure 1. Among them only the values $9, 10, 8, 9, 9, 8$ belong to the interval $[10(1-\tau), 10]$, and consequently only the first, second, fourth, fifth, sixth and seventh peak maximal occurrences are extreme. Now assume a period threshold $\delta = 2$. Since the minimum/maximum distance between any j and $j+1$ positions of extreme pattern occurrences is equal to $2/6$, the correspond range $6-2$ is greater than $\delta = 2$. Consequently the corresponding sequence of extreme peaks is not quasi-periodic of order 1. But it is quasi-periodic of order 2 because (1) the minimum/maximum distance between any j and $j+2$ positions is equal to $8/10$, and (2) the first (resp. last) occurrence of peak is not too far from the leftmost (resp. rightmost) border.

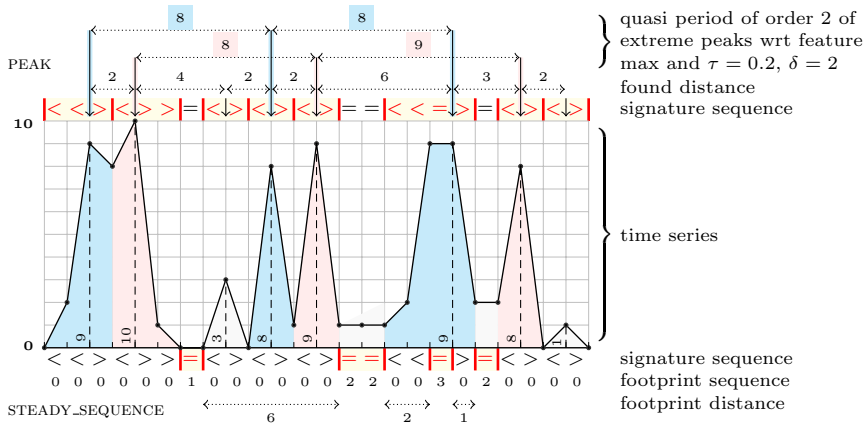


Fig. 1: Illustrating the notions of signature, found distance, footprint distance, extreme patterns and quasi-periodic extreme patterns wrt the PEAK=' $<(=\mid<)^*(>\mid=)^*>$ ' and STEADY_SEQUENCE=' $=^+$ ' patterns

3.3 CP-based time series annotation

In all what follows, a positive (respectively negative) workload trace or time series is a workload trace or time series that do not present any anomaly (respectively that is abnormal). The purpose is to label a set of historical traces as being

either positive or negative. As a first step, we begin with a small set of time series that are prelabelled as positives. From those sample time series, we identify a set $E_p = p_1, \dots p_n$ of quasi-periodic extreme patterns and we compute their respective periods. This process is presented in Algorithm 1, where:

- procedure *compute_extreme_occurrences*(s_t, f, p, e_t) uses Definition 4 to computes the list of all extreme occurrences of the pattern p with respect to the feature f and the threshold e_t.
- procedure *compute_period*(EV, p_t, g) uses Definition 9 to check if the extreme occurrences in the list EV are quasi-periodic of order g. If yes computes and return the period, else it returns the empty set.

Algorithm 1 Algorithm to acquire a CP model out of positive time series

PROCEDURE $LearnCPModel(S, P, F, e_t, p_t) : (Model)$
S: Positive time series set,
P: Set of patterns to consider,
F: Set of features to consider,
$e_t \in [0, 1]$: Extreme value threshold,
$p_t \in \mathbb{N}$: Period threshold.

$st \leftarrow \{\}$
$Model \leftarrow \{\}$
for all $s \in S$ **do**
    $st \leftarrow st \bigoplus s$ ▷ Concatenate all positive time series into st
end for
$\overline{g} \leftarrow$max order to consider
$Periods \leftarrow \emptyset$
for all $(f, p) \in F \times P$ **do**
    for all $g \in [1, \overline{g}]$ **do**
        $EV \leftarrow compute\_extreme\_occurrences(s_t, f, p, e_t)$
        $Period_{f,p,g} \leftarrow compute\_period(EV, p_t, g)$
        $Periods \leftarrow Periods \bigcup Period_{f,p,g}$
    end for
    $Model \leftarrow Model \bigcup (f, p, Periods)$
end for
return $Model$

Each of the remaining traces is labelled as positive if it presents all the quasi-periodic extreme patterns of P, with compatible periods (periods whose intersection is non-empty) i.e. if it satisfies the CP model acquired by Algorithm 1. The traces that do not satisfy the constraints of the model are labelled as negatives. This process is described in Algorithm 2. The next step trains a model that can detect anomalies at run time. The next section details the specificities of our model.

Algorithm 2 Using an acquired CP model to label time series

PROCEDURE $LabelTimeSeries(st, M, e_t, pt) : (ts, label)$

```
st:          Time series to label,
M:           Learned model,
e_t ∈ [0,1]: Extreme value threshold,
p_t ∈ ℕ:     Period threshold.

for all [(f, p, Periods)] ∈ M × P do
    EV ← compute_extreme_occurrences(s_t, f, p, e_t)
    for all Period_{f,p,g} ∈ Periods do
        TSPeriod_{f,p,g} ← compute_period(EV, p_t, g)
        if ¬IsCompatible(TSPeriod_{f,p,g}, Period_{f,p,g}) then
            label ← negative
            return (s_t, label)
        end if
    end for
end for
label ← positive
return (s_t, label)
```

4 Anomalies detection

Given a workload trace, an anomaly is said to occur iff a quasi-periodic extreme pattern, identified by the CP model acquisition step, does not occur at an expected location. In Sections 4.1 and 4.2, we propose two distinct approaches to predict those anomalies. The first one relies on a constrain programming model as a pre-training step while the second one relies on k-means clustering.

4.1 Supervised learning method

Supervised learning [9] is a class of machine learning algorithms that aim at learning a function that computes an output from an input observation. In order to compute that function, supervised learning algorithms rely on input examples together with the expected output value.

In this section we show how we use a deep neural network to detect malfunctions in a cloud infrastructure.

Neural network model There is no state-of-the-art method to set the parameters of a neural network [16]. This section presents all the empiric choices relative to our model.

Deep neural network architecture

– Input layer. The choice here is intuitive. Our workload traces are made up of time series representing daily workloads with a time step of 10 minutes. It means that each workload trace is a time series of 144 values, we thus chose 144 as the number of neurons in the input layers.

– Hidden layers. We empirically chose 2 hidden layers of 10 neurons each.
– Output layer with two neurons.
  Having two neurons in the output layer, the results are interpreted in the following way. The first neuron's value is the probability that we predict the first category i.e. *abnormal*, and the second neuron's value is the probability that we predict the second category *normal*.

Optimisation

– Activate function. We chose *softmax* as activation function since we need to output probabilistic values between two discriminative classes [16].
– Stochastic gradient descent. The output y of a neural network given an input X is given by Equation 1 where f is the activate function, W is the weight matrix and B is the bias matrix [16].

$$y = f(WX + B) \tag{1}$$

  To set the parameters W and B during the learning, we used the standard stochastic gradient descent technique with randomly shuffling the training examples. As shown in [7], it performs better than sequential selection of examples that are grouped by classes as it is the case in our model.

4.2 Unsupervised learning method

Unsupervised leaning [12] is a machine learning technique that consists in inferring a discriminative function out of an unlabelled set of data.

One of the most popular and yet simple unsupervised learning algorithm is *k-means* [14]. *k-means* partitions a set of unlabelled data into k partitions according to a given metric. One difficulty in using *k-means* is to choose k the number of clusters, as there is no mathematical criterion [19].

In the context of our application, we did not encounter that difficulty, as we already know, that we want to discriminate between two categories, *normal* and *abnormal*. Therefore we set $k = 2$ and we use the Euclidean distance as the distance metric.

The steps required to use the k-means clustering algorithm to predict an anomaly in a time series ts are the following:

1. Learning step: The purpose is to partition the dataset into two clusters.
   – Select a known normal time series ts_n and a known abnormal one ts_a.
   – Set ts_a and ts_n to be the initial centroids for k-means clustering.
   – Cl_a and Cl_n are the two clusters obtained from the dataset, such that $ts_a \in Cl_a$ and $ts_n \in CL_n$
2. Prediction step:
   – Computes the centroids c_a and c_n of Cl_a and Cl_n respectively.
   – ts is normal (positive) if $d(ts, cn) \leq d(ts, ca)$ and abnormal otherwise, where $d(x, y)$ is the Euclidean distance between the points x and y.

5 Experiments

In this section, we evaluate each step of our method to predict malfunction in a data centre, we also show the necessity of reinforcing the learned constraint programming model with the neural network of Section 4.1. We further evaluate the unsupervised based learning of Section 4.2.

5.1 Dataset

The initial input to our framework is a set of historical workload traces provided by a French SME company specialised in virtualised data centre analysis. Figure 2 shows an instance of raw historical workload traces. Each curve of the figure represents the CPU workload of a VM in a data centre for a whole day in the interval 23^{rd} of February 2017 to 26^{th} of October 2017.

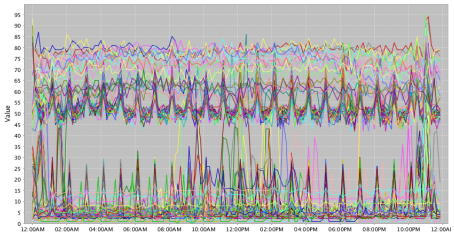


Fig. 2: Unlabeled historical data

5.2 Implementation of the CP based labelling

We implemented Algorithm 2 of Section 3.3 with SICStus Prolog [8].

Figures 3 and 4 shows how the initial unlabelled dataset of Figure 2 is partitioned into two distinct subsets using Algorithm 2 of Section 3.3.

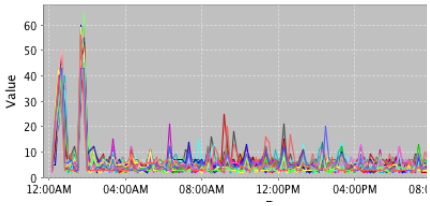


Fig. 3: Workloads automatically labelled as normal by the CP model

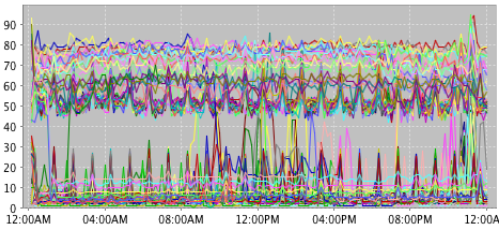


Fig. 4: Workloads automatically labelled as abnormal by the CP model

From Figure 3 we can observe that the quasi-periodic pattern acquired by the CP model is the occurrence of two extreme peaks every day between midnight and 3AM. However, as we will see in Section 5.3, we should take care of some false negative annotations.

5.3 Neural network reinforcement of the CP model

The constraint programming model learned using Algorithm 1 can be used with Algorithm 2 to label workload traces, i.e. it can also be used to detect malfunctions in a given workload. However, the quality of the CP model strongly rely on the threshold $\tau \in [0, 1]$ of Definition 3. This threshold sets the magnitude above and below which a feature value is considered extreme.

Figure 5 shows that the choice of value of $\tau \in [0, 1]$ is a trade-off between the percentage of false positive and false negative predicted by the model.

Figure 5 induces the following observations:

- For a threshold value above 0.4, i.e. for $\tau \in [0.4, 1]$ the percentage of false positive predictions of the models grows exponentially. This is not suitable as a false positive prediction is when the model fails to predict a malfunction.
- On the other $\tau \in [0, 0.4]$, the percentage of false negative prediction is close to 100. This is also not suitable as it means that the model predicts a malfunction in almost every workload.

The best trade-off threshold value is thus 0.4. Although $\tau = 0.4$ implies a perfect 0% false positive predictions, it also implies around 12% false negative predictions. This is the reason why we use a neural network on top of the CP model. The CP model provides a good learning base for the neural network.

5.4 Neural network versus K-means prediction

The purpose of the neural network in our framework is to capture hidden structures within the dataset. We implemented the model presented at Section 4.1 in Python using Tensorflow [1]. The amount of false positive and false negative predictions using the so trained neural network are both 0%.

Figure 6 shows how the precision of the model gets high over the learning steps. The precision is evaluated with the cross entropy loss function.We can see that the loss rapidly converges to value 0.

On the other hand, the prediction using the k-means method suffers from the fact that it is based on a notion of closeness through Euclidean distance. The dataset depicted at Figure 2 is therefore separated in two partitions. The direct consequence is the amount of false positive prediction that is above 70%. While a model with a reasonable false negative prediction rate may remain exploitable, the inverse is not desirable. A false negative prediction in our context means that the model fails to predict a malfunction. A 70% false positive rate makes the k-means model not exploitable in this context.

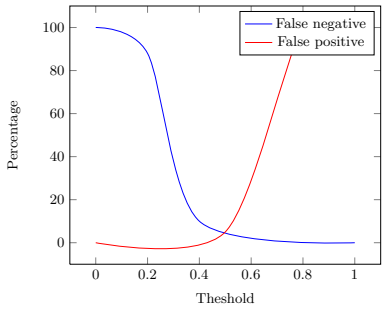


Fig. 5: False positive against false negative trade-off of the CP model

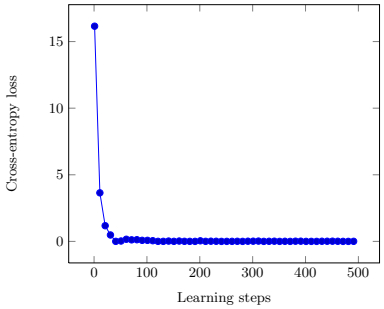


Fig. 6: Learning graph of the neural network model

6 Conclusion

In this paper we first introduced the generic notions of *extreme patterns* and *quasi-periodic pattern group* to capture the behaviour of a time series. Using those notions, we described a machine learning framework for anomaly detection in a cloud workload trace. To overcome the problem of deep unsupervised learning, we proposed a method that relies on constraint programming through the definition of quasi-periodic extreme pattern to annotate the workload time series. After acquiring a constraint programming model, the data set is automatically annotated and fed to a neural network for prediction. Experimental results on real workload traces show that state of the art unsupervised learning methods like K-means are not exploitable in our context, and that the combination of a constraint programming model and a neural network produces better results.

Acknowledgement This work was done as part of the PIA HYDDA project funded by the French government. The authors would also like to thank Frederico Alvares from EasyVirt for his active feedback.

References

1. **Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.** (**2016**). Tensorflow: A system for large-scale machine learning. In *OSDI*, volume 16. 265–283.
2. **Abbas, H., Rodionova, A., Bartocci, E., Smolka, S. A., & Grosu, R.** (**2017**). Quantitative regular expressions for arrhythmia detection algorithms. In *Int. Conf. on Computational Methods in Systems Biology*. Springer, 23–39.
3. **Alur, R., Fisman, D., & Raghothaman, M.** (**2016**). Regular programming for quantitative properties of data streams. In **Thiemann, P.**, editor, *ESOP 2016*, volume 9632 of *LNCS*. Springer, 15–40.

4. **Arafailova, E., Beldiceanu, N., Douence, R., Carlsson, M., Flener, P., Rodríguez, M. A. F., Pearson, J., & Simonis, H.** (**2016**). Global constraint catalog, volume ii, time-series constraints. *arXiv preprint arXiv:1609.08925*.
5. **Beldiceanu, N., Carlsson, M., Douence, R., & Simonis, H.** (**2016**). Using finite transducers for describing and synthesising structural time-series constraints. *Constraints*, 21(1), 22–40.
6. **Beldiceanu, N., Ifrim, G., Lenoir, A., & Simonis, H.** (**2013**). Describing and generating solutions for the EDF unit commitment problem with the modelseeker. In **Schulte, C.**, editor, *CP 2013*, volume 8124 of *LNCS*. Springer, 733–748.
7. **Bottou, L.** (**2012**). Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade*. Springer, 421–436.
8. **Carlsson, M., Widen, J., Andersson, J., Andersson, S., Boortz, K., Nilsson, H., & Sjöland, T.** (**1988**). *SICStus Prolog user's manual*, volume 3. Swedish Institute of Computer Science Kista, Sweden.
9. **Caruana, R. & Niculescu-Mizil, A.** (**2006**). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*. ACM, 161–168.
10. **Coates, A. & Ng, A. Y.** (**2012**). *Learning Feature Representations with K-Means*. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-35289-8, 561–580. doi:10.1007/978-3-642-35289-8_30.
11. **Francisco Rodríguez, M. A., Flener, P., & Pearson, J.** (**2017**). Automatic generation of descriptions of time-series constraints. In **Virvou, M.**, editor, *ICTAI 2017*. IEEE Computer Society.
12. **Hastie, T., Tibshirani, R., & Friedman, J.** (**2009**). Unsupervised learning. In *The elements of statistical learning*. Springer, 485–585.
13. **Ipeirotis, P. G., Provost, F., & Wang, J.** (**2010**). Quality management on Amazon Mechanical Turk. In *Proceedings of the ACM SIGKDD workshop on human computation*. ACM, 64–67.
14. **Jain, A. K.** (**2010**). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8), 651–666.
15. **Madi-Wamba, G., Li, Y., Orgerie, A., Beldiceanu, N., & Menaud, J.** (**2017**). Green energy aware scheduling problem in virtualized datacenters. In *ICPADS 2017*. IEEE Computer Society, 648–655.
16. **Nielsen, M. A.** (**2015**). *Neural networks and deep learning*. Determination Press.
17. **Paolacci, G., Chandler, J., & Ipeirotis, P. G.** (**2010**). Running experiments on amazon mechanical turk.
18. **Scholkmann, F., Boss, J., & Wolf, M.** (**2012**). An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals. *Algorithms*, 5(4), 588–603.
19. **Tibshirani, R., Walther, G., & Hastie, T.** (**2001**). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411–423.
20. **Turk, A. M.** (**2012**). Amazon Mechanical Turk. *Retrieved August*, 17, 2012.
21. **Vishwanath, K. V. & Nagappan, N.** (**2010**). Characterizing cloud computing hardware reliability. In *Proceedings of the 1st ACM symposium on Cloud computing*. ACM, 193–204.

Time series modelling with MATLAB: the SSpace toolbox

Diego J. Pedregal*, Marco Villegas*, Diego Villegas*, and Juan R. Trapero**

*Industrial Engineering School
**Faculty of Chemical Science and Technology
Universidad de Castilla-La Mancha, Ciudad Real, Spain

1 Abstract

SSpace is a powerful and flexible MATLAB toolbox that provides a number of advanced and up-to-date routines designed for a general analysis of State Space (SS) systems. The toolbox is rather different to other pieces of software, though it takes ideas from many sources, like **CAPTAIN** [1], **SSfPack** [2, 3], **E4** [4] and others. A partial review of alternatives is available in volume 41, 2011 of the Journal of Statistical Software [5].

Regarding statistical issues, the main properties of **SSpace** are:

1. The particular SS formulation implemented is very general at least in three aspects: i) full multivariate linear and non-linear Gaussian models, and univariate non-Gaussian models are implementable; ii) non-linear, time varying or transfer function relations with inputs are possible; and iii) all system matrices are potentially time-varying or state-dependent.
2. Three recursive algorithms are available, namely the Kalman Filter, Fixed Interval Smoother and Disturbance Smoother. All three are implemented with exact, diffuse or ad-hoc initialisation. Exact initialisation of non-linear models is also implemeted following [6].
3. Whenever possible, **SSpace** automatically takes advantage of exact scores in Maximum Likelihood estimation, otherwise a numerical approximation is used. The user may choose numerical scores whenever he/she considers appropriate.
4. Apart from Maximum Likelihood in time domain, other estimation procedures are implemented. At the moment the toolbox offers Concentrated Likelihood and minimization of combinations of several steps ahead forecast errors. More cost functions may appear in the future.
5. Univariate treatment of multivariate systems is possible via a particular implementation of the Kalman filtering and smoothing algorithms due to [3], pages 155-160.
6. Univariate and multivariate versions of Dynamic Harmonic Regression models are implemented. As long as the authors are concerned, this is the first time that a multivariate Seemingly Unrelated Dynamic Harmonic Regression is used and implemented. Such a model is an extension of the univariate Dynamic Harmonic Regression counterpart [7].

In addition, **SSpace** is rather flexible and user friendly because:

- The general design of the toolbox is at the service of the final user because a big programming effort has been done to make its usage as simple as possible. A comprehensive time series analysis may be done with a reduced number of functions that follows a simple and fixed structure easy to remember. Despite this simplicity, it is possible to have full control over models, parameters and specifications. For example, just one function (namely SSfilter) is used to filter any system, whether it is univariate, multivariate, time-varying, linear, non-linear or non-gaussian. This is possible because the toolbox detects automatically the type of system and uses the appropriate algorithms without any intervention from the user.
- Names of all functions have been chosen very carefully following nemotechnic rules, so that the final user may remember them or may look for them easily. There are three groups of functions:
  - Core functions (named SS*) for setting up models and perform the basic operations, like filtering, forecasting, smoothing, validating, etc.
  - Template functions (Sample*) that allows the user to set up any model either from scratch (from the low level State Space form) or from a middle level, i.e. from typical model structures, like regression, transfer functions, ARIMA, ExponenTial Smoothing, Unobserved Components, etc.
  - Other helpful functions for an easy handling of models, matrices, etc. The most common are those for differencing time series, building forecast confidence bands, constraining parameters, building semi-definite covariance matrices, etc.
- Models are directly specified by the user in a user-coded function written in plain **MATLAB** syntax. This approach has at least the following advantages:
  - The syntax to handle all time series rely on the coded function and is fixed and independent of the type of system, i.e. the same syntax applies to all system types and the same functions are used for system operations. Internally such operations may be rather different, but no extra intervention from the user is required.
  - Templates are supplied to aid the user with building different types of SS systems, that include a wide range of standard models. The catalog may be modified or extended by the user.
  - Additional templates are provided for performing important operations, like concatenating and nesting SS systems or specifying systems with time aggregation constraints.
  - Absolute control over the systems is possible. For example, arbitrary linear or non-linear parameter constraints of any kind may be imposed upon the models, different specific parameterizations of the same model are possible, non-standard features of the models may be added (like heteroskedasticity, time varying parameters, non linear eXogenous relations, ...), etc.

– Different optimisers may be used to estimate the models. The default option is fminunc.m from the standard optimisation toolbox in **MATLAB**, but the particular options to this funciton or other optimisation routins may be used by editing the script optimizer.m.

The code is freely available on-line at `https://bitbucket.org/predilab/sspace-matlab/`, together with a manual that helps the potential user to learn how to use the toolbox step by step. The power of **SSpace** is demonstrated by 8 demos, some of them standard, but some others rather unusual.

References

1. Taylor, C., Pedregal, D.J., Young, P.C., Tych, W.: Environmental tiem series analysis and forecasting with the Captain toolbox. Environmental Modelling & Software, 22, 797-814 (2007)
2. Koopman, S.J., Shephard, N., Doornik, J.A.: Statistical Algorithms for Models in State Space Form: **SsfPack 3.0**. Timberlake Consultants Press (2008)
3. Durbin, J and Koopman, SJ: Time series analysis by state space methods. Oxford University Press (2012)
4. Casals, J.M., Garcia-Hiernaux, A., Jerez, M., Sotoca, S., Trindade, A.A.: State-Space Methods for Time Series Analysis: Theory, Applications and Software. Chapman and Hall/CRC (2016)
5. Commandeur, J.J.F., Koopman, S.J., Ooms, M.: Statistical Software for State Space Methods. Journal of Statistical Software. 41 (2011)
6. Koopman, S.J., Lee, K.M.: Seasonality with trend and cycle interactions in unobserved components models. Journal of the Royal Statistical Society C (Applied Statistics). 58, 427–448 (2009)
7. Young, P.C., Pedregal, D.J., Tych, W.: Dynamic harmonic regression. Journal of Forecasting. 18, 369–394 (1999)

Multivariate INAR processes - Periodic case

Cláudia Santos[a,b], Isabel Pereira[b], and Manuel Scotto[c]

[a]Coimbra College of Agriculture, Polytechnic Institute of Coimbra
[b]CIDMA, University of Aveiro
[c]IST & CEMAT, University of Lisbon
{csps@ua.pt}

Abstract. *A multivariate integer-valued autoregressive model of order one with periodic time-varying parameters, and driven by a periodic innovations sequence of independent random vectors is established. The binomial thinning operator replaces the scalar multiplication in the common time series models. The matricial form of the multivariate model and its basic statistical properties are defined. Emphasis is placed upon models with periodic multivariate negative binomial innovations. Aiming to reduce computational burden arising from the use of the conditional maximum likelihood method a composite likelihood-based approach is adopted and compared with other traditional competitors.*

Keywords: Multivariate models, binomial thinning operator, composite likelihood

1 Introduction

Multivariate count data can occur in many fields. Special attention has been devoted to bivariate integer-valued time series processes (e.g. Pedeli and Karlis [6], [8]) based on the binomial thinning operator introduced by Steutal and van Harn ([9]). The role of the innovations is significant since not only they determine the joint distribution of the two series under consideration but also they form the unique source of cross-correlation. The assumption of a bivariate negative binomial distribution for the innovations allows for more flexibility than the Poisson `BINAR`(1) model due to the involvement of the overdispersion parameter.

Within the reasonably large spectrum of integer-valued models proposed in the literature only a few focus on the modeling of multivariate time series of count data with periodic structure. Our interest in periodic integer-valued autoregressive models was primarily influenced by the work of Monteiro *et al.* whose periodic univariate and bivariate `INAR` models were introduced in [4] and in [5], respectively. We seek to extend `INAR` models to multi-dimensional space, assuming periodic time-varying parameters and periodic sequences of innovations. Apart from the general specification of such models, we also examined their statistical properties and proposed alternative estimation techniques.

2 Periodic MINAR model of order one

Let $\{\mathbf{X}_t\}$ be a periodic m-variate integer-valued autoregressive process of first-order defined by the recursion

$$\mathbf{X}_t = A_t \circ \mathbf{X}_{t-1} + \mathbf{Z}_t,\ t \in \mathbb{Z}, \tag{1}$$

where $\mathbf{X}_t, \mathbf{X}_{t-1}$ and $\mathbf{Z}_t$ are random ms-vectors with $\mathbf{X}_t = [\mathbf{X}_{1,t}\ \mathbf{X}_{2,t}\ \cdots\ \mathbf{X}_{m,t}]'$ for $t = v + ns$, $v = 1, \ldots, s$ and $n \in \mathbb{N}_0$, and $\mathbf{X}_{j,t} = [X_{j,1+ns}\ \cdots\ X_{j,s+ns}]'$, $j = 1, \ldots, m$. The ms-dimensional vector $\mathbf{Z}_t = [\mathbf{Z}_{1,t}\ \mathbf{Z}_{2,t}\ \cdots\ \mathbf{Z}_{m,t}]'$ constitutes a periodic sequence of independent random vectors with

$$\mathbf{Z}_{j,t} = [Z_{j,1+ns}\ Z_{j,2+ns}\ \cdots\ Z_{j,s+ns}]'. \tag{2}$$

The model in (1) will be referred to as the Periodic Multivariate INteger-valued AutoRegressive model of order one (PMINAR(1) in short) with period $s \in \mathbb{N}$. The PMINAR(1) model admits the following matricial representation

$$\begin{bmatrix} \mathbf{X}_{1,t} \\ \mathbf{X}_{2,t} \\ \vdots \\ \mathbf{X}_{m,t} \end{bmatrix} = \begin{bmatrix} \phi_{1,t} & 0 & \cdots & 0 \\ 0 & \phi_{2,t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{m,t} \end{bmatrix} \circ \begin{bmatrix} \mathbf{X}_{1,t-1} \\ \mathbf{X}_{2,t-1} \\ \vdots \\ \mathbf{X}_{m,t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \vdots \\ \mathbf{Z}_{m,t} \end{bmatrix}, \tag{3}$$

where matrix A_t in (1) is a $(ms \times ms)$-diagonal matrix, representing the periodic integer-valued autoregressive coefficients in season v and $\phi_{j,t} = \alpha_{j,v} \in (0,1)$ for $t = v + ns; v = 1, \ldots, s; n \in \mathbb{N}_0$ and $j = 1, \ldots, m$.

For each t, $\mathbf{Z}_{j,t}$ is assumed to be independent of $\mathbf{X}_{j,t-1}$ and $\phi_{j,t} \circ \mathbf{X}_{j,t-1}$. Note that the j-th component in (3) is

$$\mathbf{X}_{j,t} = \phi_{j,t} \circ \mathbf{X}_{j,t-1} + \mathbf{Z}_{j,t}, \tag{4}$$

with $\phi_{j,t} \circ X_{j,t-1} \overset{d}{=} \sum_{r=1}^{X_{j,t-1}} U_{r,t}(\phi_{j,t})$, where $\{U_{r,t}(\phi_{j,t})\}_{r\in\mathbb{N}}$ is a periodic sequence of i.i.d. Bernoulli-distributed random variables with probability of success $P(U_{r,t}(\phi_{j,t}) = 1) = \phi_{j,t}$. Since the autocorrelation matrix A_t is diagonal, the only source of dependence between the series $\mathbf{X}_{1,t}, \ldots, \mathbf{X}_{m,t}$ in (3) is given through $\mathbf{Z}_t$. Therefore, the innovations will play a central role in the specification of the PMINAR(1) process.

Due to the fact that $t = v + ns$, then $\mathbf{X}_{j,t-s} \overset{d}{=} \mathbf{X}_{j,v+(n-1)s}$ $(v = 1, \ldots, s)$, meaning that the j-th component $\mathbf{X}_{j,t}$ in (4) can be expressed as

$$\mathbf{X}_{j,t} = A_j \circ \mathbf{X}_{j,t-s} + B_j \circ \mathbf{Z}_{j,t}, \tag{5}$$

where the $(s \times s)$-matrices A_j and B_j $(j = 1, \ldots, m)$ are given by

$$A_j = \begin{bmatrix} 0 & \cdots & 0 & \alpha_{j,1} \\ 0 & \cdots & 0 & \alpha_{j,2}\alpha_{j,1} \\ 0 & \cdots & 0 & \alpha_{j,3}\alpha_{j,2}\alpha_{j,1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \prod_{k=0}^{s-1} \alpha_{j,s-k} \end{bmatrix} \tag{6}$$

and

$$B_j = \begin{bmatrix} 1 & 0 & 0 & \dots 0 \\ \alpha_{j,2} & 1 & 0 & \dots 0 \\ \alpha_{j,3}\alpha_{j,2} & \alpha_{j,3} & 1 & \dots 0 \\ \vdots & \vdots & \vdots & \ddots \vdots \\ \prod_{k=0}^{s-2} \alpha_{j,s-k} & \prod_{k=0}^{s-3} \alpha_{j,s-k} & \prod_{k=0}^{s-4} \alpha_{j,s-k} & \dots 1 \end{bmatrix}, \tag{7}$$

respectively, with coefficients $\alpha_{j,v} \in (0,1)$, $j = 1, \dots, m$ and $v = 1, \dots, s$. Taking all m components, the PMINAR(1) model defined in (1) can be rewritten in the form

$$\mathbf{X}_t = \widetilde{A} \circ \mathbf{X}_{t-s} + \widetilde{B} \circ \mathbf{Z}_t, \tag{8}$$

The $(ms \times ms)$-matrices $\widetilde{A}$ and $\widetilde{B}$ in equation (8) are block-diagonal matrices, that is

$$\widetilde{A} = diag(A_1, A_2, \dots, A_m) \tag{9}$$

and

$$\widetilde{B} = diag(B_1, B_2, \dots, B_m) \tag{10}$$

with matrices A_j and B_j $(j = 1, \dots, m)$ as in (6) and in (7), respectively. Generally, matrix $\widetilde{A}$ has entries a_{ik}^j satisfying $0 \leq a_{ik}^j < 1$ and matrix $\widetilde{B}$ has entries b_{ik}^j satisfying $0 \leq b_{ik}^j \leq 1$ with $i, k = 1, \dots, ms$ and $j = 1, \dots, m$. Furthermore, it will be assumed that the innovations $\mathbf{Z}_t$ have finite first- and second-order moments taking the form

$$E[\mathbf{Z}_t] = E \begin{bmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \\ \vdots \\ \mathbf{Z}_{m,t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\delta}_{1,t} \\ \boldsymbol{\delta}_{2,t} \\ \vdots \\ \boldsymbol{\delta}_{m,t} \end{bmatrix} =: \boldsymbol{\delta}_t. \tag{11}$$

The ms-mean vector $\boldsymbol{\delta}_t$ with $t = v + ns; v = 1, \dots, s$ and $n \in \mathbb{N}_0$ has m $(s \times 1)$-vectors, i.e.,

$$E[\mathbf{Z}_{j,t}] = \boldsymbol{\delta}_{j,t} = \begin{bmatrix} \lambda_{j,1} & \lambda_{j,2} & \dots & \lambda_{j,s} \end{bmatrix}', \tag{12}$$

for $j = 1, \dots, m$. For a fixed v, each element in (12) is

$$E[Z_{j,v+ns}] = \lambda_{j,v}. \tag{13}$$

Turning to the variance-covariance matrix of $\mathbf{Z}_t$, it follows that

$$\sum\nolimits_{\mathbf{Z}_t} = \begin{bmatrix} Var[\mathbf{Z}_{1,t}] & Cov(\mathbf{Z}_{1,t}, \mathbf{Z}_{2,t}) & \cdots & Cov(\mathbf{Z}_{1,t}, \mathbf{Z}_{m,t}) \\ & Var[\mathbf{Z}_{2,t}] & \cdots & Cov(\mathbf{Z}_{2,t}, \mathbf{Z}_{m,t}) \\ & & \ddots & \vdots \\ & & & Var[\mathbf{Z}_{m,t}] \end{bmatrix} = \begin{bmatrix} \psi_{11,t} & \psi_{12,t} & \dots & \psi_{1m,t} \\ & \psi_{22,t} & \dots & \psi_{2m,t} \\ & & \ddots & \vdots \\ & & & \psi_{mm,t} \end{bmatrix} =: \psi_t, \tag{14}$$

where $\psi_{jk,t}$ $(j, k = 1, \dots, m; t = v + ns; v = 1, \dots, s; n \in \mathbb{N}_0)$ are $(s \times s)$- diagonal matrices of the form

$$\psi_{jk,t} = diag(\sigma_{jk,1}, \dots, \sigma_{jk,s}), \tag{15}$$

and for a fixed v, each element of the diagonal in matrix (15) is given by

$$\sigma_{jk,v} = Cov(Z_{j,v+ns}, Z_{k,v+ns}). \tag{16}$$

2.1 Strictly periodically stationary distribution

Let PMINAR(1) be the process in (8). The existence of a periodically stationary solution to (8) depends on the largest eigenvalue of the non-negative matrix $\widetilde{A}$ in (9), whose coefficients $\alpha_{j,v} \in (0,1)$ for all components.

Lemma 1 For a fixed v $(v = 1, \ldots, s)$, $\alpha_{j,v} \in (0,1)$ where $j = 1, \ldots, m$ and for $t = v + ns$, $0 < P(\mathbf{Z}_t = \underline{0}) < 1$. Furthermore, any solution to $\{\mathbf{X}_t\}$, $t = v + ns$ and $n \in \mathbb{N}_0$ in (8) is an irreducible and aperiodic Markov chain.

The main result of this subsection is formalized through the theorem below.

Theorem 1 For a fixed v $(v = 1, \ldots, s)$, let $\{\mathbf{X}_t\}$ with $t = v + ns$ and $n \in \mathbb{N}_0$ as in (8) be an irreducible, aperiodic Markov chain on $\mathbb{N}_0^m$. If $E||\mathbf{Z}_t|| < +\infty$ and if the largest eigenvalue of $\widetilde{A}$ is less than one, then there exists a strictly periodically stationary (or cyclostationary) m-variate INAR(1) process satisfying recursion (8).

Proof. From Lemma 1, $\{\mathbf{X}_t\}$ with $t = v + ns$ and fixed $v = 1, \ldots, s$ is an irreducible and aperiodic Markov chain being the eigenvalues of matrix $\widetilde{A}$ are less than one. Thus by Franke and Subba Rao ([2]) a strictly periodically stationary m-variate non-negative integer-valued process satisfying the equation (8) exists.

The PMINAR(1) model in (8) can be expressed as

$$\mathbf{X}_t = \widetilde{A} \circ \mathbf{X}_{t-s} + \mathbf{R}_t, \tag{17}$$

where $\mathbf{R}_t = \widetilde{B} \circ \mathbf{Z}_t$ with matrix $\widetilde{B}$ in (10).

2.2 Mean vector of cyclostationary PMINAR(1)

In this subsection we obtain the periodic mean and autocovariance function of the PMINAR(1) model. First note that the expectation of $\mathbf{R}_t$ is

$$E[\mathbf{R}_t] = E[\widetilde{B} \circ \mathbf{Z}_t] = \widetilde{B} E[\mathbf{Z}_t] = \widetilde{B}\boldsymbol{\delta}_t \tag{18}$$

with matrices $\widetilde{B}$ and $\boldsymbol{\delta}_t$ as in (10) and (11), respectively. Furthermore, for each component $j = 1, \ldots, m$, the mean vector of $\mathbf{R}_{j,t}$ takes the form

$$E[\mathbf{R}_{j,t}] = \begin{bmatrix} \lambda_{j,1} \\ \lambda_{j,1}\alpha_{j,2} + \lambda_{j,2} \\ \lambda_{j,1}\alpha_{j,3}\alpha_{j,2} + \lambda_{j,2}\alpha_{j,3} + \lambda_{j,3} \\ \vdots \\ \lambda_{j,1}\prod_{k=0}^{s-2}\alpha_{j,s-k} + \lambda_{j,2}\prod_{k=0}^{s-3}\alpha_{j,s-k} + \cdots + \lambda_{j,s-1}\alpha_{j,s} + \lambda_{j,s} \end{bmatrix}.$$

Moreover,

$$\boldsymbol{\mu}_t = E[\mathbf{X}_t] = E[\widetilde{A} \circ \mathbf{X}_{t-s} + \mathbf{R}_t] = (I - \widetilde{A})^{-1}\widetilde{B}\boldsymbol{\delta}_t \tag{19}$$

with matrices $\widetilde{A}$ and $\widetilde{B}$, and vector $\boldsymbol{\delta}_t$ as in (9), (10) and (11), respectively. Note that the ms-dimensional mean vector $\boldsymbol{\mu}_t$ for $t = v + ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$ in (19) takes the form $\boldsymbol{\mu}_t = \left[\boldsymbol{\mu}_{1,t}\ \boldsymbol{\mu}_{2,t}\ \ldots\ \boldsymbol{\mu}_{m,t}\right]'$.

For each $j = 1, \ldots, m$ and $l \geq i$, let

$$\varphi_{l,i}^{(j)} = \begin{cases} \prod_{k=0}^{i-1} \alpha_{j,l-k},\ i \geq 1 \\ 1, \qquad i = 0 \end{cases}. \tag{20}$$

It follows by tedious (although straightforward) calculations that for a fixed v and j, each entry in $\boldsymbol{\mu}_{j,t} = [E(X_{j,1+ns})\ E(X_{j,2+ns}) \cdots E(X_{j,s+ns})]'$ is given by

$$E(X_{j,v+ns}) = \frac{\sum_{k=0}^{v-1} \varphi_{v,k}^{(j)} \lambda_{j,v-k} + \varphi_{v,v}^{(j)} \sum_{i=0}^{s-(v+1)} \varphi_{s,i}^{(j)} \lambda_{j,s-i}}{1 - \varphi_{s,s}^{(j)}}, \tag{21}$$

for $j = 1, \ldots, m;\ t = v + ns;\ v = 1, \ldots, s$ and $n \in \mathbb{N}_0$. We adopt the convention $\sum_{i=0}^{s-(s+1)} \varphi_{s,i}^{(j)} \lambda_{j,s-i} = 0$.

2.3 Variance-covariance matrix

To derive the variance-covariance matrix of $\{\mathbf{X}_t\}$ we start by calculation the variance-covariance matrix, $\sum_{\mathbf{R}_t}$, of $\{\mathbf{R}_t\}$. To this extent note that,

$$\sum\nolimits_{\mathbf{R}_t} = \widetilde{B} \sum\nolimits_{\mathbf{Z}_t} \widetilde{B}' + diag(Q\boldsymbol{\delta}_t) = \widetilde{B}\psi_t\widetilde{B}' + diag(Q\boldsymbol{\delta}_t)$$

with matrices $\widetilde{B}$, $\boldsymbol{\delta}_t$ and ψ_t in (10), (11) and (14), respectively. Moreover the $(ms \times ms)$-matrix $Q = \widetilde{B}(I - \widetilde{B})$ has entries $[q_{ik}^j] = [b_{ik}^j(1 - b_{ik}^j)]_{i,k=1,\ldots,ms}$ for $j = 1, \ldots, m$, being b_{ik}^j the elements of matrix $\widetilde{B}$ in (10). Thus, matrix Q is also block-diagonal with m $(s \times s)$-matrices Q_j.

For simplicity in notation we define

$$\sum\nolimits_{\mathbf{X}_t} =: \begin{bmatrix} \sum_{1,1} & \sum_{1,2} & \cdots & \sum_{1,m} \\ & \sum_{2,2} & \cdots & \sum_{2,m} \\ & & \ddots & \vdots \\ & & & \sum_{m,m} \end{bmatrix}.$$

Note that

$$\sum\nolimits_{j,j} = \begin{bmatrix} Var[X_{j,1+ns}] & Cov(X_{j,1+ns}, X_{j,2+ns}) & \ldots & Cov(X_{j,1+ns}, X_{j,s+ns}) \\ & Var[X_{j,2+ns}] & \ldots & Cov(X_{j,2+ns}, X_{j,s+ns}) \\ & & \ddots & \vdots \\ & & & Var[X_{j,s+ns}] \end{bmatrix}$$

with diagonal elements

$$Var[X_{j,v+ns}] = \frac{g(\varphi,\lambda)}{1-\left(\varphi_{s,s}^{(j)}\right)^2}, \tag{22}$$

where

$$\begin{aligned} g(\varphi,\lambda) := &\sum_{k=0}^{v-1} \varphi_{s,s}^{(j)}\varphi_{v,k}^{(j)}\lambda_{j,v-k} + \varphi_{v,k}^{(j)}\left(1-\varphi_{v,k}^{(j)}\right)\lambda_{j,v-k} + \left(\varphi_{v,k}^{(j)}\right)^2\sigma_{j,v-k}^2 + \\ &+ \sum_{m=0}^{s-(v+1)} \varphi_{s,s}^{(j)}\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\lambda_{j,s-m} + \varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\left(1-\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\right)\lambda_{j,s-m} + \\ &+ \sum_{m=0}^{s-(v+1)} \left(\varphi_{v,v}^{(j)}\varphi_{s,m}^{(j)}\right)^2\sigma_{j,s-k}^2, \end{aligned}$$

for a fixed v $(v = 1,\ldots,s)$ and off-diagonal elements

$$Cov(X_{j,v+ns}, X_{j,v+ns+l}) = \varphi_{v+l,l}^{(j)} Var[X_{j,v+ns}]. \tag{23}$$

2.4 PMINAR(1) process with MVNB Innovations

In this subsection we derive the first- and second-order moment structure of the PMINAR(1) process driven by periodic multivariate negative binomial (MVBN) innovations. Hence, the joint probability mass function

$$\begin{aligned} &h(z_1,\ldots,z_m) = P(Z_{1,v+ns}=z_1,\ldots,Z_{m,v+ns}=z_m) = \\ &= \frac{\Gamma\left(\beta_v^{-1}+\sum_{j=1}^m z_j\right)}{\Gamma(\beta_v^{-1})}\left(\frac{\beta_v^{-1}}{\beta_v^{-1}+\sum_{j=1}^m\lambda_{j,v}}\right)^{\beta_v^{-1}}\left(\beta_v^{-1}+\sum_{j=1}^m\lambda_{j,v}\right)^{-\sum_{j=1}^m z_j}\cdot \\ &\cdot\prod_{j=1}^m\frac{\lambda_{j,v}^{z_j}}{z_j!},\ (z_1,\ldots,z_m)\in\mathbb{N}_0^m. \end{aligned} \tag{24}$$

Notice the marginal distribution of $\mathbf{Z}_{j,t}$ is univariate negative binomial with parameters β_v^{-1} and $p_{j,v}$ $(j = 1,\ldots,m; v = 1,\ldots,s)$ with

$$p_{j,v} = \frac{\beta_v^{-1}}{\lambda_{j,v}+\beta_v^{-1}}. \tag{25}$$

The innovation process $\{\mathbf{Z}_t\}$, $t = v+ns;\ v = 1,\ldots,s$ and $n \in \mathbb{N}_0$ is generally defined as a periodic sequence of independent random vectors with mean as in (11) and variance-covariance matrix as in (14), respectively. Thus,

$$\lambda_{j,v} = E[Z_{j,v+ns}] = \beta_v^{-1}\ \frac{1-p_{j,v}}{p_{j,v}}, \tag{26}$$

$$\sigma_{j,v}^2 = Var[Z_{j,v+ns}] = \beta_v^{-1}\ \frac{1-p_{j,v}}{p_{j,v}^2} = \lambda_{j,v}(1+\beta_v\lambda_{j,v}), \tag{27}$$

$$\sigma_{jk,v} = Cov(Z_{j,v+ns}, Z_{k,v+ns}) = \beta_v\lambda_{j,v}\lambda_{k,v}, \tag{28}$$

for a fixed v ($v = 1, \ldots, s$), $j \neq k$; $j, k = 1, \ldots, m$. Note that $Var[Z_{j,v+ns}] > E[Z_{j,v+ns}]$. Thus, the first-order moment and the auto- and cross-covariance structure PMINAR(1) process are obtained from (21), (22) and (23) by plugging-in the values of $\lambda_{j,v}$, $\sigma^2_{j,v}$ and $\sigma_{jk,v}$ as in (26)-(28).

3 Parameter estimation

Consider a finite time series $\{\mathbf{X}_{j,ts}\}$ with $1 \leq t \leq N$, $j = 1, \ldots, m$ (N-number of complete cycles) from the PMINAR(1) model in (17) with MVNB innovations. Without loss of generality it will be assumed that $\mathbf{X}_0 = \mathbf{x}_0$. The vector of unknown parameters $\boldsymbol{\theta}$ is a $(2m+1)s$-dimensional vector

$$\boldsymbol{\theta} := (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \boldsymbol{\beta}) \tag{29}$$

with s-vectors ($j = 1, \ldots, m$):

$$\boldsymbol{\alpha}_j = (\alpha_{j,1}, \ldots, \alpha_{j,s}) \; ; \; \boldsymbol{\lambda}_j = (\lambda_{j,1}, \ldots, \lambda_{j,s}) \; ; \; \boldsymbol{\beta} = (\beta_1, \ldots, \beta_s). \tag{30}$$

In order to estimate the unknown parameters in $\boldsymbol{\theta}$, three estimation methods are proposed, namely: Yule-Walker, conditional maximum likelihood and composite likelihood.

3.1 Yule-Walker estimation

The YW estimator of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}_{YW} := (\widehat{\boldsymbol{\alpha}}_1^{YW}, \ldots, \widehat{\boldsymbol{\alpha}}_m^{YW}, \widehat{\boldsymbol{\lambda}}_1^{YW}, \ldots, \widehat{\boldsymbol{\lambda}}_m^{YW}, \widehat{\boldsymbol{\beta}}^{YW})$ are calculated as follows: first, the YW estimators of parameters $\boldsymbol{\lambda}_j$ are calculated through the solution of the system of s linear equations yielding

$$\widehat{\lambda}_{j,v}^{YW} = \begin{cases} \overline{X}_{j,v} - \widehat{\alpha}_{j,v}^{YW} \overline{X}_{j,s} & , v = 1 \\ \overline{X}_{j,v} - \widehat{\alpha}_{j,v}^{YW} \overline{X}_{j,v-1} & , v = 2, 3, \ldots, s \end{cases}, \tag{31}$$

where $\overline{X}_{j,v} = \frac{1}{N} \sum_{n=0}^{N-1} X_{j,v+ns}$, $j = 1, \ldots, m$, is the sample mean. The YW estimators of parameters $\boldsymbol{\alpha}_j$ are given by

$$\widehat{\alpha}_{j,v}^{YW} = \begin{cases} \frac{S^2_{j,v}}{S^2_{j,s}} & , v = 1 \\ \frac{\gamma_{j,v-1}(1)}{S^2_{j,v-1}} & , v = 2, 3, \ldots, s \end{cases}, \tag{32}$$

where $S^2_{j,v}$ is the sample variance and $\gamma_{j,v}(1)$ ($j = 1, \ldots, m$) is defined as

$$\gamma_{j,v}(1) = Cov(X_{j,v+ns}, X_{j,v+1+ns}) =$$
$$= \begin{cases} \frac{1}{N-1} \sum_{n=0}^{N-1} (X_{j,v+ns} - \overline{X}_{j,v})(X_{j,v+1+ns} - \overline{X}_{j,v+1}) & , v = 1, \ldots, s-1 \\ \frac{1}{N-1} \sum_{n=0}^{N-1} (X_{j,v+ns} - \overline{X}_{j,v})(X_{j,1+(n+1)s} - \overline{X}^*_{j,1}) & , v = s \end{cases},$$

with $\overline{X}^*_{j,1} = \frac{1}{N}\sum_{n=0}^{N} X_{j,1+ns}$. Finally, the YW estimator of β_v is

$$\widehat{\beta}_v^{YW} = \frac{\left(1 - \widehat{\varphi}^{(j)}_{s,s}\widehat{\varphi}^{(k)}_{s,s}\right)\gamma_{jk,v}(0)}{\sum_{i=0}^{v-1}\widehat{\varphi}^{(j)}_{v,i}\widehat{\varphi}^{(k)}_{v,i}\widehat{\lambda}_{j,v-i}\widehat{\lambda}_{k,v-i} + \widehat{\varphi}^{(j)}_{v,v}\widehat{\varphi}^{(k)}_{v,v}\sum_{i=0}^{s-(v+1)}\widehat{\varphi}^{(j)}_{s,i}\widehat{\varphi}^{(k)}_{s,i}\widehat{\lambda}_{j,s-i}\widehat{\lambda}_{k,s-i}}, \quad (33)$$

for $v = 1, \ldots, s$ and $j, k = 1, \ldots, m$ $(j \neq k)$.

3.2 Conditional maximum likelihood estimation

In order to obtain the CML estimator $\widehat{\boldsymbol{\theta}}_{CML}$ of $\boldsymbol{\theta}$ we proceed as follows. First note that the transition probabilities for the `PMINAR`(1) model can be expressed as the convolution of m binomials with parameters $(x_{j,v-1+ns}, \alpha_{j,v})$ for $v = 1, \ldots, s$ with probability mass function

$$f_j(r_j) = C^{x_{j,v-1+ns}}_{r_j}\alpha^{r_j}_{j,v}(1-\alpha_{j,v})^{x_{j,v-1+ns}-r_j},\ j = 1, \ldots, m, \quad (34)$$

and the periodic discrete m-variate distribution defined as in (24). Thus, the conditional density is the multiple sum

$$p_v(\mathbf{x}_{v+ns}|\mathbf{x}_{v-1+ns}) = P(\mathbf{X}_{v+ns} = \mathbf{x}_{v+ns}|\mathbf{X}_{v-1+ns} = \mathbf{x}_{v-1+ns}) =$$
$$= \sum_{r_1=0}^{g_1}\sum_{r_2=0}^{g_2}\cdots\sum_{r_m=0}^{g_m}\left(\prod_{j=1}^{m} f_j(r_j)\right)h(x_{1,v+ns}-r_1, \ldots, x_{m,v+ns}-r_m) \quad (35)$$

with $g_j := min(x_{j,v-1+ns}, x_{j,v+ns})$. Hence, the CML estimator is obtained by maximizing the conditional log-likelihood $C(\boldsymbol{\theta}) = ln(L(\boldsymbol{\theta}|\mathbf{x}))$. Explicit CML estimators are not available so numerical procedures have to be employed. The asymptotic properties of $\widehat{\boldsymbol{\theta}}_{CML}$ are given through the following result.

Theorem 2 The conditional maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{CML}$ of $\boldsymbol{\theta}$ is asymptotically normal

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, I^{-1}(\boldsymbol{\theta}))$$

where $I(\boldsymbol{\theta})$ represents the Fisher information matrix, i.e., $I = diag(M_1, \ldots, M_s)$ with matrices M_v $(v = 1, \ldots, s)$ given by

$$(-1)\times\begin{bmatrix}
E\left[\frac{\partial^2 C(\theta)}{\partial\alpha^2_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}\partial\alpha_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}\partial\lambda_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}\partial\lambda_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{1,v}\partial\beta_v}\right] \\
\vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\
E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}\partial\alpha_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha^2_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}\partial\lambda_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}\partial\lambda_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\alpha_{m,v}\partial\beta_v}\right] \\
E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}\partial\alpha_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}\partial\alpha_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda^2_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}\partial\lambda_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{1,v}\partial\beta_v}\right] \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}\partial\alpha_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}\partial\alpha_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}\partial\lambda_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda^2_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\lambda_{m,v}\partial\beta_v}\right] \\
E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v\partial\alpha_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v\partial\alpha_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v\partial\lambda_{1,v}}\right] & \cdots & E\left[\frac{\partial^2 C(\theta)}{\partial\beta_v\partial\lambda_{m,v}}\right] & E\left[\frac{\partial^2 C(\theta)}{\partial\beta^2_v}\right]
\end{bmatrix}.$$

Proof. This result is a particular case of Theorem 2.2 in [1]. For each v ($v = 1, \ldots, s$), $p_v(\cdot|\cdot)$ is the transition probabilities in (35) of the PMINAR(1) model, therefore the regularity conditions in Billingsley's Theorem 2.2 are satisfied.

3.3 Composite likelihood estimation

For periodic multivariate processes, the number of parameters can be quite large due to season v ($v = 1, \ldots, s$) with s representing the period. Computational issues often arise when applying the conditional maximum likelihood approach, the complexity of the method augments with dimensional increase. Composite likelihood inherits many of the good properties of inference based on the full likelihood function, but is more easily implemented with high-dimensional data sets. These methods based on optimizing sums of log-likelihoods of low-dimensional margins have become popular in recent years (e.g. [7]); being useful for multivariate models in which the likelihood of multivariate data is very time-consuming. The methodology has drawn considerable attention in a broad range of applied disciplines in which complex data structures arise (e.g. [10]).

Note that the bivariate marginal log-likelihood function between two random elements, say X_a and X_b, can be defined as

$$l_{ab}(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b) = \frac{1}{Ns} \sum_{n=0}^{N-1} \sum_{v=1}^{s} log f_{X_a,X_b}(x_{a,v+ns}, x_{b,v+ns} | x_{a,v-1+ns}, x_{b,v-1+ns}; \boldsymbol{\theta}), \tag{36}$$

where

$$\begin{aligned} & f_{X_a,X_b}(x_{a,v+ns}, x_{b,v+ns} | x_{a,v-1+ns}, x_{b,v-1+ns}; \boldsymbol{\theta}) = \\ & = \sum_{k_a=0}^{g_1} \sum_{k_b=0}^{g_2} \binom{x_{a,v-1+ns}}{x_{a,v+ns}-k_a} \alpha_{a,v}^{x_{a,v+ns}-k_a} (1-\alpha_{a,v})^{x_{a,v-1+ns}-x_{a,v+ns}+k_a} \cdot \\ & \cdot \binom{x_{b,v-1+ns}}{x_{b,v+ns}-k_a} \alpha_{b,v}^{x_{b,v+ns}-k_b} (1-\alpha_{b,v})^{x_{b,v-1+ns}-x_{b,v+ns}+k_b} \cdot h_{R_a,R_b}(k_a, k_b) \end{aligned}$$

for $g_1 = min(x_{a,v+ns}, x_{a,v-1+ns})$ and $g_2 = min(x_{b,v+ns}, x_{b,v-1+ns})$. The bivariate function $h_{R_a,R_b}(k_a, k_b)$ represents the bivariate marginal probability density with bivariate negative binomial innovations being a particular case ($m = 2$) of the multivariate negative binomial distribution in (24).

The composite log-likelihood function, $cl(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b)$, then arises as the sum of all bivariate log-likelihood functions, i.e.,

$$cl(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b) = \sum_{a=1}^{m-1} \sum_{b=a+1}^{m} w_{ab}\ l_{ab}(\boldsymbol{\theta}; \mathbf{x}_a, \mathbf{x}_b), \tag{37}$$

where w_{ab} is a constant weight for l_{ab}. For sake of simplicity, it is common to set $w_{ab} = 1$, $1 \leq a \leq b \leq m$ (e.g. [3]).

4 Simulation study

The performance of the three estimation methods of the `PMINAR`(1) model driven by multivariate negative binomial innovations is compared through a simulation experiment for $m = 3$ (trivariate). We have set period $s = 4$ therefore $\boldsymbol{\theta} := (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\beta})$. This study contemplates the following set of parameters: $\boldsymbol{\alpha}_1 = (0.53, 0.75, 0.62, 0.83)$, $\boldsymbol{\alpha}_2 = (0.72, 0.85, 0.56, 0.91)$, $\boldsymbol{\alpha}_3 = (0.83, 0.60, 0.41, 0.58)$ and $\boldsymbol{\lambda}_1 = (4, 2, 3, 5)$, $\boldsymbol{\lambda}_2 = (5, 3, 1.2, 2)$, $\boldsymbol{\lambda}_3 = (3, 1.6, 2, 4)$ and $\boldsymbol{\beta} = (1.6, 0.9, 1.8, 1.2)$. Three alternative samples sizes where considered: $n = 400, 1000, 2000$. Thus $n = sN$, $N = 100, 250, 500$ complete cycles. For each experiment we conducted 200 independent replications.

Comparison of the YW, CML and CL estimators was made in terms of the mean square error and the biases of the produced estimates. Table 1 reports the estimates for autocorrelation parameters $\boldsymbol{\alpha}_j$ $(j = 1, 2, 3)$, where small MSE characterize all estimates of $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$.

Table 1. YW, CML and CL estimates for $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3}, \alpha_{j,4})$ with $j = 1, 2, 3$. Mean square error in parenthesis.

	n =400			n =1000			n =2000		
	YW	CML	CL	YW	CML	CL	YW	CML	CL
$\hat{\alpha}_{1,1}$	0.521	0.531	0.531	0.528	0.531	0.531	0.528	0.529	0.529
	(0.0018)	(5.1×10^{-7})	(0.0002)	(0.0001)	(0.00002)	(0.00004)	(0.0001)	(0.00002)	(0.00003)
$\hat{\alpha}_{1,2}$	0.746	0.752	0.751	0.750	0.751	0.751	0.752	0.749	0.748
	(0.0001)	(0.00005)	(0.0008)	(0.0006)	(0.0009)	(0.0009)	(0.0001)	(0.00002)	(0.00003)
$\hat{\alpha}_{1,3}$	0.608	0.618	0.617	0.617	0.621	0.620	0.615	0.620	0.619
	(0.0004)	(0.0026)	(0.00002)	(0.0074)	(0.0011)	(0.0013)	(0.0005)	(0.00006)	(0.00006)
$\hat{\alpha}_{1,4}$	0.789	0.833	0.832	0.826	0.830	0.830	0.825	0.830	0.830
	(0.0111)	(0.0007)	(0.0006)	(0.0020)	(0.00002)	(0.00001)	(0.0011)	(0.00007)	(0.00006)
$\hat{\alpha}_{2,1}$	0.717	0.718	0.717	0.739	0.719	0.719	0.740	0.720	0.720
	(0.0027)	(0.00002)	(0.0038)	(0.0001)	(0.00003)	(0.00002)	(0.0010)	(7.5×10^{-6})	(0.00006)
$\hat{\alpha}_{2,2}$	0.845	0.854	0.852	0.845	0.851	0.851	0.849	0.851	0.850
	(0.0001)	(0.0008)	(0.00003)	(0.0002)	(0.00002)	(0.00002)	(0.0004)	(0.00003)	(0.00003)
$\hat{\alpha}_{2,3}$	0.552	0.559	0.560	0.559	0.559	0.559	0.561	0.560	0.560
	(0.0002)	(0.0025)	(0.00009)	(0.0006)	(0.00007)	(0.00007)	(0.0002)	(0.00004)	(0.00002)
$\hat{\alpha}_{2,4}$	0.894	0.910	0.910	0.910	0.910	0.911	0.906	0.909	0.910
	(0.0105)	(0.0001)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0001)	(0.00001)	(0.00002)
$\hat{\alpha}_{3,1}$	0.823	0.832	0.832	0.832	0.831	0.831	0.830	0.830	0.830
	(0.0071)	(0.0005)	(0.0013)	(0.0001)	(0.00006)	(0.0001)	(0.0003)	(0.00001)	(0.00003)
$\hat{\alpha}_{3,2}$	0.596	0.603	0.603	0.601	0.600	0.600	0.599	0.601	0.602
	(0.0011)	(0.0008)	(0.0020)	(0.0020)	(0.00002)	(0.0004)	(0.0001)	(0.0003)	(0.0004)
$\hat{\alpha}_{3,3}$	0.391	0.411	0.410	0.411	0.409	0.409	0.407	0.411	0.411
	(0.0002)	(0.0001)	(0.0009)	(0.0004)	(0.0037)	(0.0040)	(0.0001)	(8.2×10^{-6})	(0.00002)
$\hat{\alpha}_{3,4}$	0.545	0.587	0.588	0.566	0.580	0.580	0.578	0.580	0.580
	(0.0002)	(0.0020)	(0.0046)	(0.0155)	(0.0003)	(0.0003)	(0.0028)	(0.0027)	(0.0028)

The tendency of the YW method to produce inadmissible estimates was greater for smaller sample sizes. YW estimates were used as initial values in numerical routines for the optimization procedure of CML and CL methods. The performance of the estimators $\hat{\boldsymbol{\lambda}}_j$ $(j = 1, 2, 3)$ and estimator $\hat{\boldsymbol{\beta}}$ (not shown here) is slightly worse. The estimates obtained by adopting either the CML or the CL method are very close to the real parameter values, even in the case of a moderate sample size ($n = 400$). For larger samples ($n = 1000, 2000$), both estimators seem to perform well and in a similar way.

Graphical inspection is provided through the boxplots of the biases of the produced estimates. Fig. 1 displays the boxplots of the biases of the estimates for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$, the dispersion parameter, where the effect of sample size on the behavior of the estimators can be seen. As expected, increasing the sample size improves the performance of all estimators in terms of both location (median closer to zero) and dispersion (narrower interquartile ranges). Small and not definite differences are observed between CML and CL methods, regarding both location and dispersion. Therefore, this indicates the superiority of CML and CL estimators over the YW estimator.

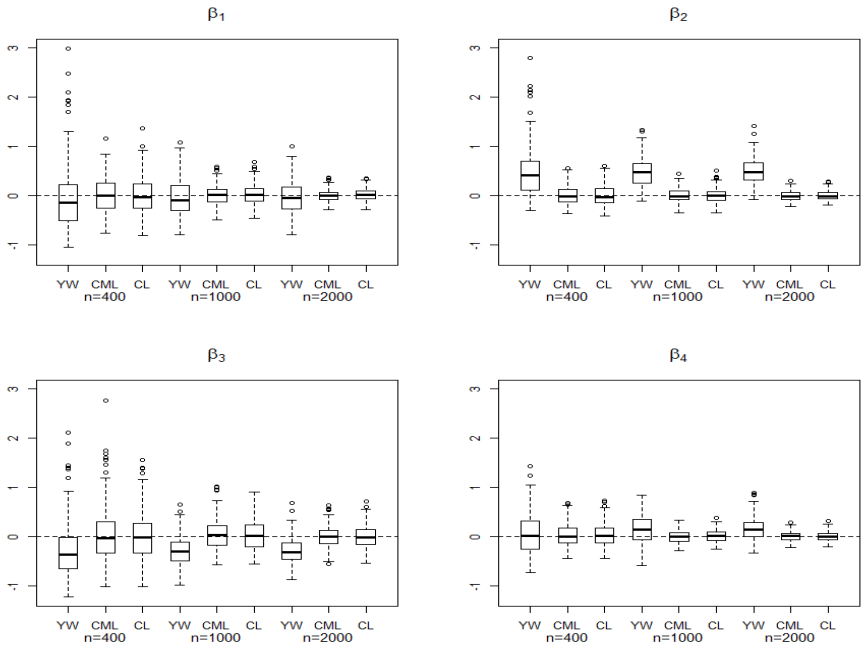


Fig. 1. Boxplots for the biases of the YW, CML and CL estimates of the parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$. From left to right, the first three boxplots display the biases of $\hat{\beta}_1$ for the three methods with $n = 400, 1000, 2000$. The same information follows for $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$, respectively.

Acknowledgments. Portuguese Foundation for Science and Technology (FCT-Fundacão para a Ciência e a Tecnologia) through CIDMA-Center for Research and Development in Mathematics and Applications (University of Aveiro), within project UID/MAT/04106/2013.

References

1. Billingsley, P.: Statistical Inference for Markov Processes. Statistical Research Monographs, University of Chicago Press (1961)
2. Franke, J., Subba Rao, T.: Multivariate first–order integer–valued autoregression, Technical report, Universität Kaiserslautern (1993)
3. Joe, H., Lee, Y.: On weighting of bivariate margins in pairwise likelihood. J. Mult. Anal. 100, 670–685 (2009)
4. Monteiro, M., Scotto, M.G., Pereira, I.: Integer–valued autoregressive processes with periodic structure. J. Stat. Plan. Inf. 140, 1529–1541 (2010)
5. Monteiro, M., Scotto, M.G., Pereira, I.: A periodic bivariate integer-valued autoregressive model. In: Bourguignon J.P., Jelstch R., Pinto A., Viana M. (eds). Dynamics, Games and Science - International Conference and Advanced School Planet Earth, DGS II, pp. 455–477. Springer International Publishing (2015)
6. Pedeli X., Karlis, D.: A bivariate INAR(1) process with application. Stat. Modelling 11, 325–349 (2011)
7. Pedeli, X., Karlis, D.: On composite likelihood estimation of a multivariate INAR(1) model. J. Time S. Anal. 34, 206–220 (2013a)
8. Pedeli, X., Karlis, D.: On estimation of the bivariate Poisson INAR process. Commun. in Stat: Simulation and Comput. 42, 514–533 (2013b)
9. Steutel, F.W., van Harn, K.: Discrete analogues of self–decomposability and stability. The Annals of Prob. 7, 893–899 (1979)
10. Varin, C., Reid, N.M., Firth, D.: An overview of composite likelihood methods. Stat. Sinica 21, 5–42 (2011)

The Impact of Feedback Trading on Option Prices

Thorsten Lehnert*

Luxembourg School of Finance, University of Luxembourg, Luxembourg

March 2018

Extended Abstract

This paper examines whether S&P 500 index option prices are affected by feedback trading. Our testing framework is a heterogeneous agent's option pricing model, where agents differ in their level of rationality. Hence, they form different beliefs about the future level of market volatility, and trade options accordingly. We introduce feedback traders, who incorporate noisy signals into their volatility beliefs. We find that feedback trading appears to be an important determinant of index option prices. In addition, we find that feedback trading partly explains the term structure of the index option smile, because it primarily affects short-term options.

We assume that traders in the options market can be classified into three different groups that have different expectations about the future evolution of index volatility: fundamentalists, who trade on the principle of mean reversion; chartists, who trade on exogenous shocks; and feedback traders, who trade on noisy signals from the mutual fund market. Typically, stochastic volatility option pricing models are implanted using Monte Carlo simulation techniques. We propose a new method to calibrate the model based on the Filtered Historical Simulation approach. We make use of the fact that we can match the average fund flows – our indicator for noise trader sentiment' – with the standardized residuals for all days that we use to estimate the empirical distribution. Therefore, in the calibration of the models by sampling from the empirical distribution, we obtain not only the historical standardized residual, but pairs of the news innovations and the noise signals. The method has the implicit advantage that we pick up the exact historical correlation between returns and fund flows without the need to explicitly model them as a separate process.

In our empirical application, we calibrate all models on each Wednesday of the years 2005, 2006 and 2007, and calculate the RMSE's in-sample as well for one week out-of-sample. We find that calibrating the model with a nonparametric innovation distribution compared to standard Monte Carlo techniques results in slightly lower pricing errors. When we incorporate noise trading in the pricing model, we recognize that chartists and noise traders incorporate their information into their beliefs about volatility in a similar way. They expect volatility to go down for good news (positive returns or fund flows) and to go up for bad news (negative returns or fund flows). However, once we include noise traders' trades, in-sample as well as out-of-sample pricing errors of the models are reduced. Given that we only make minor modifications to the specifications, the differences are sizeable. Results further suggest that on average 25% of traders follow a fundamentalist strategy, 19% a chartists strategy and, as a result, about 56% a noise traders strategy. Additionally, noise traders react more strongly to fund flows for shorter maturities compared to longer maturities, but long-term investors are more sensitive to differences in forecasting performance between the chartists and noise trader strategies compared to short-term traders.

Hence, we present an interesting alternative to the well-known Monte Carlo simulation techniques and investigate the benefits in an option pricing exercise. We show that noise trader risk is an important determinant of prices and that noise trading can partly explain the volatility dynamics underlying option prices.

Keywords: Feedback Trading, Option Valuation, Heterogeneous Agents, Filtered Historical Simulations.

JEL-Classification: G12, C15

* Luxembourg School of Finance, University of Luxembourg, 4, rue Albert Borschette, L-1246 Luxembourg, Luxembourg, tel +352466644-6941; fax +352466644-6835. E-mail address: thorsten.lehnert@uni.lu .

Physical Laws Extracted from Statistical Analyses of Solar Magnetic Elements

Mohsen Javaherian, PhD in Solar and Astrophysics,
ZNU, IRAN

Hossein Safari, Professor in Solar and Astrophysics,
ZNU, IRAN

Abstract (300 word limit)

Studying the emergence and disappearance of magnetic features on the solar surface is controversial. Two approaches were introduced to investigate the evolution of photospheric features. One way is magnetohydrodynamics, and the other is statistical analyses. The complexity of this system with lots of features proposes the statistical way can be adopted to carry out this kind of nonlinear systems.

The solar photosphere is the origin of magnetic elements observed in the visible and UV wavelengths. Using Helioseismic and Magnetic Imager (HMI) on board Solar Dynamic Observatory (SDO) telescope, the magnetic field of photosphere is measured in line-of-sight. Magnetic elements (patches) can be appeared on magnetograms. The code Yet Another Feature Tracking Algorithm (YAFTA). First, an area 400"×400" was picked up to study the magnetic characteristics over the year 2011. The correlation between filling factors of negative and positive polarities is 0.51. The relationship between the size [S] and flux [F] of elements has a power-law behavior as $S \propto F^{0.69}$. Next, the area 100"×100" is selected from solar equatorial region to track patches in the quiet-Sun over three days with time lags of 45 seconds. It was included 5750 sequential frames to investigate lifetime of elements and its relationship with the flux and size. The code YAFTA detected 22,526 patches survived for more than one frame. More than about 95% of the magnetic features have lifetimes less than 100 minutes, and approximately 0.05% of the detected features had lifetimes of more than 6 hours with maximum lifetimes of 15 hours.

The relationships between S, lifetime [T], and F for patches bring in power-law behaviors as $S \propto T^{0.25}$ and $F \propto T^{0.38}$, respectively. The DFA analysis suggests there is long-range memories in the system of patches which can give an information about the conditions govern on patches.

Image

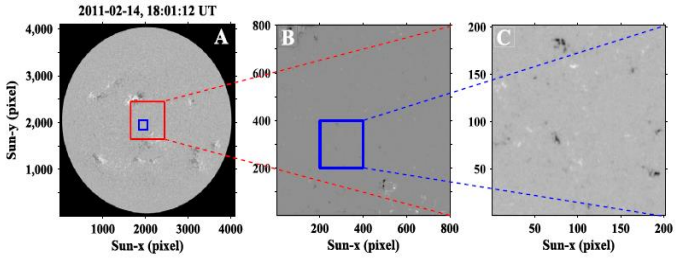


Figure 1 SDO/HMI full-disk magnetogram of the Sun recorded on 14 February 2011 (18:01:12 UT) (**a**). The cutout image with an area of 400 × 400 arcsec^2 (**b**) from the solar equatorial region is selected to extract physical parameters of positive and negative elements (such as size distributions, flux distributions, and filling factors) in 2011 with a cadence of one image per day. An image tile with an area of 100 × 100 arcsec^2 (**c**) is cropped from all of the images to investigate the physical parameters of the magnetic elements, especially their lifetimes for three days (14 – 16 February 2011) in the QS, with a time lag of 45 seconds between frames.

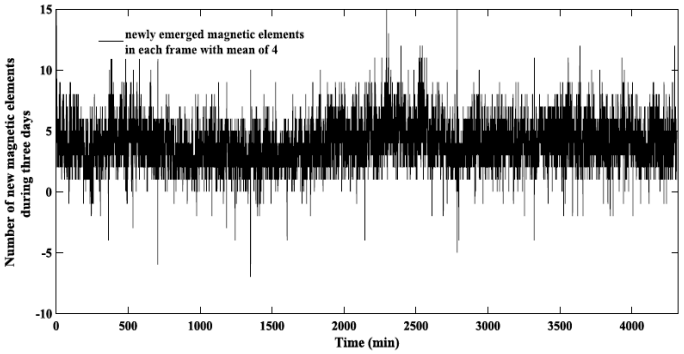


Figure 2 Time series of new magnetic elements that appeared in each magnetogram for the three-day dataset with a cadence of 45 seconds. The average number of newly emerged magnetic elements in each frame is equal to four elements. The value of the Hurst exponent extracted with the DFA method is equal to 0.82.

Recent Publications (minimum 5)

1. Gosain S (2012) Evidence for Collapsing Fields in the Corona and Photosphere during the 2011 February 15 X2.2 Flare: SDO/AIA and HMI Observations. Astrophys. J. 749: 85-94.
2. DeForest D E et al. (2007) Solar Magnetic Tracking. I. Software Comparison and Recommended Practices. Astrophys. J. 666: 676-587.
3. Javaherian M et al. (2017) Statistical Properties of Photospheric Magnetic Elements Observed by HMI onboard SDO. Solar Phys. 292: 1-15.
4. de Wijn A G et al. (2009) Small-Scale Solar Magnetic Fields. Space Sci. Rev. 144: 275-315.

A robust alternative for the estimation of autocovariance from the frequency domain for multivariate processes

Higor Cotta[1,2], Valdério Reisen[1,2], Pascal Bondon[2], and Céline Lévy-Leduc[3]

[1]NuMEs - DEST/PGGEA - Federal University of Espírito Santo - Brazil
[2]L2S - CentraleSupélec - France
[3]MMIP - AgroParisTech
higor.cotta@l2s.centralesupelec.fr
valderio.reisen@ufes.br
pascal.bondon@l2s.centralesupelec.fr
celine.levy-leduc@agroparistech.fr

Abstract. This paper presents a new method for the robust estimation of the sample cross-covariance and cross-correlation functions in the presence of additive outliers. This robust methodology is achieved by replacing the standard Fourier transform with its robust version, i.,e, obtained by substituting the least square procedure in the harmonic regression by the non-linear M-regression. Simulation experiments are conducted to assess the performance of the estimators under contaminated and non-contaminated scenarios.

Keywords: Cross-covariance function, additive outliers, M-cross-periodogram, estimation

In the time series literature, it is well-known outliers can lead to the complete destruction of the correlation structure and, thus, leading to the misspecification of the model and wrong conclusions. This issue is found for univariate and multivariate time series. However, as pointed by [1], most of the outliers studies are devoted only to the univariate time series or to the multivariate time uncorrelated processes. As tangible evidence, the autocorrelation and cross-correlation functions, for univariate and multivariate time series, respectively, are entirely destroyed by additive outliers, see, for example, [2, 3] and the references therein.

In order to address this issue, [4] studied the impact of additive outliers on the autocovariance and autocorrelation matrix function and proposed a multivariate robust estimation procedure for these functions as a way to mitigate the effects of outlying observations. The proposed estimators are based on the univariate case proposed by [5] which make use of the Q_n scale estimator proposed by [6] and further studied by [7].

The robust autocovariance functions of [5] and [4] are not positive definite. Therefore, considering a different approach, [8] proposed new estimators for the sample autocovariance and autocorrelation functions of univariate time series.

These estimators are based on the robust M-periodogram studied in [9] for short and long-memory processes.

Therefore, in this paper, we extend to multivariate stationary time series the robust estimator of the autocovariance and the autocorrelation functions proposed by [8] for univariate stationary time series from the frequency domain. The robust sample estimators are obtained from the robust M-cross-periodogram. That is, the robust M-cross-periodogram is achieved calculating the Fourier coefficients from the robust M-regression. Then, the multivariate ACOVF and ACF are obtained from an inverse procedure to diagonalize the matrix containing the robustly estimated multivariate spectral density.

References

1. Tsay, R.S., Peña, D., Pankratz, A.E.: Outliers in multivariate time series. Biometrika **87**(4) (2000) 789–804
2. Chan, W.: Outliers and financial time series modelling: a cautionary note. Mathematics and Computers in Simulation **39**(3) (1995) 425–430
3. Molinares, F.F., Reisen, V.A., Cribari-Neto, F.: Robust estimation in long-memory processes under additive outliers. Journal of Statistical Planning and Inference **139**(8) (2009) 2511–2525
4. Cotta, H., Reisen, V., Bondon, P., Stummer, W.: Robust estimation of covariance and correlation functions of a stationary multivariate process. In: Proc. Int. Conf. Time Series, Granada. (2017)
5. Ma, Y., Genton, M.G.: Highly robust estimation of the autocovariance function. Journal of Time Series Analysis **21** (2000) 663–684
6. Rousseeuw, P.J., Croux, C.: Alternatives to the median absolute deviation. Journal of the American Statistical Association **88**(424) (1993) 1273–1283
7. Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M.S., Reisen, V.A.: Robust estimation of the scale and of the autocovariance function of Gaussian short-and long-range dependent processes. Journal of Time Series Analysis **32**(2) (2011) 135–156
8. Cotta, H., A. Reisen, V., Bondon, P.: Robust autocovariance estimation from the frequency domain. In: International Work-Conference on Time Series, Granada, Spain (September 2017) 1073–1074
9. Reisen, V., Lévy-Leduc, C., Cotta, H., Bondon, P., Ispany, M.: An overview of robust spectral estimators,. In: submitted

Changes in rapeseed canopy spectral reflectance under different cultivars and nitrogen levels

CAO Hong-xin[1], CHEN Wei-tao[2], ZHANG Bao-jun[2], ZHANG Wen-yu[1], ZHANG Wei-Xin[1], GE Dao-ke[1], SONG Chu-wei[3], GE Si-jun[4], ZHANG Qian[5], XIA Ji-an[6], FENG Chun-huan[3], ZHOU Dong[1], WAN Qian[3]

([1]Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences/Engineering Research Center for Digital Agriculture, Nanjing 210014, China; [2]Northwest Agriculture and Forestry University, Yangling 712100, Shaanxi, China; [3]College of Agronomy，Nanjing Agricultural University, Nanjing 210095, China; [4]College of Agronomy，Yanzhou University, Yangzhou 225127, China; [5]College of Resources＆ Environment，Nanjing Agricultural University，Nanjing 210095, China; [6]Nanjing University of Science and Technology, Nanjing 210094, China)

It is particularly important that obtain and diagnosis crop growing, nitrogen nutrition status real-timely, fast and non-destructively through informatization technology. This experiments were carried out at Jiangsu Provincial Academy of Agricultural Experimental Farm from 2014 to 2016, the field experiments for varieties and fertilizer were set up. Using of Crop Scan MS16R, and ASD Field Spec Hand-Held, the spectrum reflectance for various rapeseed growth period was measured. Through rapeseed spectral information, physiological and biochemical indices and statistical analysis, the relationships between the spectral reflectance characteristics and the leaf nitrogen contents for rape seedlings under the various cultivars and nitrogen levels had been cleared, the sensitive spectral bands in rape seedlings leaf nitrogen content were found, and the rapeseed monitoring model of leaf nitrogen content for rapeseed seedlings were constructed based on spectrum. This provided a theoretical basis and technical support for the use of remote sensing technology in rapeseed nitrogen application, nondestructive sensing monitoring for leaf nitrogen, and fertilizer recommendation suitability, etc. The results were as follows:

1. Changes in the spectral reflectance from transplanting to flowering stage for rapeseed gradually decreased in the visible range, while gradually increased in the near infrared region; However, after flowering, canopy spectral reflectance increased gradually with the development of postponement in the visible range, but gradually reduced in the near infrared region.

2. Rapeseed spectral reflectance had significant differences under different nitrogen levels, and three rape varieties had similar trend, especially in the

near-infrared region. The increased nitrogen levels improved spectral reflectance in the near-infrared region, but in the visible place, increased nitrogen levels reduced spectral reflectance.

3. Between the varieties of rapeseed, changes in rapeseed canopy spectral reflectance curves were the same. But there were some differences in canopy spectrum reflectance between varieties of rapeseed. It may be due to the different crop varieties, effect of background soil and weed coverage. There are some differences in the whole growth process, even under the same conditions of field management, resulting in the difference between the spectra of different varieties.

4. At 870nm and 1320 nm bands there were a very significant correlation on spectral reflectance and leaf nitrogen content in seedling phase .The coefficient of determination R^2 was 0.651 and 0.670. Through single-band linear and nonlinear regression analysis, the coefficient of determination presented regularity of the regression equation corresponding to each band, That was that: $R^2_{Polynomial} > R^2_{Logarithm} > R^2_{Linear} > R^2_{Exponentiation} > R^2_{Index}$ .R_{870} and R_{1320}. The corresponding to the polynomial regression equation between the single band reflectance and rape leaf nitrogen content, R2 reached to 0.73 and 0.795, so it can be used to characterize the quantitative relationship between 870nm and 1320 nm bands.

KEY WORDS: Rapeseed; Nitrogen; Spectrum; Changes; Vegetation indices

Application of Deep-Learning Algorithm for Inflow Series Forecasting in South Korea

Jun-Haeng Heo[1], Ju-Young Shin[1[0000-0002-1520-3965]], and Taereem Kim[1]

[1] Yonsei University, Seoul, South Korea

Abstract. Applications of the deep-learning algorithm have led to improvement and been a breakthrough of conventional issues. A number of deep-learning algorithms has been applied for modeling and forecasting time-series. It is still questionable in the amount of improvement from the deep-learning algorithm in time-series forecast, especially to the application of the deep-learning algorithm in time series of hydrological variables. The forecast of the hydrological variable series is a recurrent challenge in the hydrological field due to complexity of their generating mechanism. Particularly, seasonality and non-linear characteristics in the time series data make the forecasting difficult. In the hydrological variable series forecasting, the deep-learning algorithm is needed to be compared with the conventional time series model such as seasonal autoregressive integrated moving average (SARIMA), artificial neural network (ANN), and artificial neuro-fuzzy inference system (ANFIS) models. Thus, in this study, prediction power of the deep-learning algorithm was investigated and compared with the other conventional models. Monthly reservoir inflow series of three dams in South Korea were used for the modeling. The long-short term memory (LSTM) network model was applied as the deep-learning algorithm. For the conventional time series models, SARIMA, ANN, and ANFIS models were applied. For one-month ahead forecast, the LSTM network model outperformed other employed conventional time series models. The ANN and ANFIS models led to lower correlation coefficients than one of SARIMA. For the very long-term forecast, that more than three months, the LSTM network led to better prediction power than others although the prediction power is not good.

Keywords: Reservoir inflow forecasting, Deep-learning algorithm, Long-Short term memory network.

1 Introduction

Many deep-learning algorithms have been applied for modeling and forecasting time-series. It is still questionable in the amount of improvement from the deep-learning algorithm in time-series fore-cast, especially to the application of the deep-learning algorithm in time series of hydrological variables. The forecast of the hydrological variable series is a recurrent challenge in the hydrological field due to complexity of their generating mechanism. Particularly, seasonality and non-linear characteristics in the time series data make the forecasting difficult. In the hydrological variable series fore-casting, the deep-learning algorithm is needed to be compared with the conventional time series model such as seasonal autoregressive integrated moving aver-age (SARIMA), artificial neural network (ANN), and artificial neuro-fuzzy inference system (ANFIS) models. Thus, in this study, prediction power of the deep-learning algorithm was investigated and compared with the other conventional models. Monthly reservoir inflow series of three dams in South Korea were used for the modeling. The long-short term memory (LSTM) network model was applied as the deep-learning algorithm. For the conventional time series models, SARIMA, ANN, and ANFIS models were applied.

2 Methods

2.1 Long-short term memory network

The LSTM is a special kind of recurrent neural network (RNN). The RNN has connections between neurons and form a directed cycle. This network structure makes an internal self-looped cell that allows to present dynamic temporal behavior. A RNN composed of LSTM units is called an LSTM network. A common LSTM unit consists of a cell, an input gate, an output gate and a forget gate. The cell represents amounts or values of memory over arbitrary time intervals. Thus, in the LSTM, each of the three gates can be thought of as a "conventional" artificial neuron. The LSTM often provides good performance for classification and prediction in time series modeling because of structure of the LSTM,.

2.2 Seasonal Autoregressive Integrated Moving Average

The Box-Jenkins approach using the Autoregressive Integrated Moving Average (ARIMA) family of models is widely employed to forecast future values based on an observed time series. The ARIMA model parsimoniously interprets a stochastic process with autoregressive (AR) and moving average (MA) operators. If there is seasonality in the time series, the SARIMA model is more useful for modeling as it considers seasonality through a differencing procedure [1].

2.3 Artificial Neural Network

The ANN model is one of the machine learning techniques that is designed to mimic the structure of the brain. It has been widely applied in hydrology to improve the predictability of future hydrologic variables because it considers both linear and nonlinear structures. In general, the ANN model is comprised of three layers (input, hidden, and output) and each layer has nodes like the neurons in the brain. The nodes contain activation functions that are connected to each other with weight parameters. The weight parameters indicate the strength of the connections between the nodes.

2.4 Artificial Neuro-Fuzzy Inference System

The ANFIS model is a multilayer feed-forward network that combines the ANN model and fuzzy logic based on the Takagi–Sugeno fuzzy inference system [2]. It is a powerful tool for hydrological forecasting because it integrates the advantages of both ANN and fuzzy inference systems.

References

1. Box, G.E.P., Jenkins, G.M.: Series Analysis Forecasting and Control, First edn. Holden-Day, San Francisco (1976).
2. . Jang, J.S.R.: ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Transactions on Systems, Man and Cybermetics 23(3), 665-685 (1993).

Evaluation of Atmospheric Particulate Matter (PM_{10}) Time Series in Badajoz, 2010-2015

Selena Carretero-Peña[1], Conrado Miró Rodríguez[2*] and Eduardo Pinilla-Gil[1]

[1] Department of Analytical Chemistry and IACYS, University of Extremadura,
Av. de Elvas s/n, 06006 Badajoz, Spain
(selenacarretero@unex.es; epinilla@unex.es)
[2] Department of Applied Physics, University of Extremadura,
Av. de la Universidad s/n, 10071 Cáceres, Spain
*cmiro@unex.es

Abstract. Particulate matter (PM) is one of the most damaging air pollutants to human health, especially the low size, respirable $PM_{2.5}$ and PM_{10} fractions. For a given location, the complex combination of a range of highly variable natural and anthropogenic sources make it essential to appeal to statistical tools for a proper characterization of particulate matter behavior. In this work, we have selected the city of Badajoz (Southwest Spain) to explore the evolution of the concentration of PM_{10} in the air during the period 2010 - 2015 by the study of the daily average data time series, using the data measured by the regional air quality network. The results have shown that daily PM_{10} concentration values have a defined time series structure, whose mathematical expression for the trend allowed us to estimate that PM_{10} concentration in Badajoz has decreased linearly at a rate of 0.0025 ($\mu g \cdot m^{-3}$)/day during the studied period. The study of the seasonal nature of the time-series allowed us to determine that the concentration of PM_{10} varied cyclically each 365 ± 2 day (annual cycle). Finally, the analysis of the time series has identified anomalous values of PM_{10} concentration that could be due to natural events (notably Saharan dust outbreaks) during the studied period.

Keywords: particulate matter (PM), time-series, trend, seasonality, air quality monitoring network.

1 Introduction

Particulate Matter (PM) is the most striking form of air pollution that can be seen to the naked view, is defined as the set of solid or liquid particles in suspension in the atmosphere, altering its natural composition. Its origin are natural phenomena such as marine aerosol, volcanic eruptions, forest fires, the dust storms and other types of soil resuspension, and anthropogenic emissions such as industrial activity or traffic. Most relevant due to enhanced health and ecological impacts are the particles under 10 μm (respirable). Not only the particles itself are dangerous, but also their chemical composition so depending on the origin they can carry a wide range of pollutants such as polycyclic

aromatic hydrocarbons and heavy metals. The combination of these factors makes particulate matter one of the most dangerous air pollutants because it can promote significant adverse effects to the health of the population like respiratory, carcinogenic and cardiovascular illness and premature deaths [1].

In recent years, the study of time series of atmospheric pollutants, including PM, have become a useful statistical tool to know their behaviour in the atmosphere, allowing to make forecasts about their levels and evolution. Using large air quality databases, which span decades, time series allow studying variations of PM concentrations in different urban, industrial and rural environments. Several studies have been published aimed to know the relationship between PM levels and vehicular traffic [2], determining the weekly, seasonal and diurnal variation of PM levels and their relationship with wind speed [3], establishing correlations between PM levels and weather conditions [4] and studying the influence of PM levels on the evolution of the number of cases of some diseases [5].

In this work, we aim to initiate the study on the time pattern of the levels of PM in the Extremadura region, selecting the city of Badajoz as a model location. We have focused on the concentration of particulate matter (PM_{10} fraction) to explore its behaviour in recent times (2010-2015) and the capability of the model to predict future levels, as a tool for improving scientific knowledge and environmental protection management.

2 Materials and methods

2.1 Study area

Badajoz is a city located in the southwest of the Iberian Peninsula (Extremadura region). The city is located on the Portuguese border, on the banks of the Guadiana River. It has Mediterranean climate with Atlantic and African influence, characterized by mild winters, very hot and dry summers, irregular rainfall and sometimes persistent mists. It has a population of 150,000 inhabitants. Badajoz has an intense commercial activity, this being the dominant economic sector in the city. The city is a trade node, receiving clients from the surrounding areas in Spain and Portugal, implying a remarkable traffic of passengers and goods by road and rail. Industrial activity are however not very developed. These natural and anthropogenic characteristics condition the complex city's atmospheric environment composition and evolution, making it ideal for the application of time series statistics.

2.2 Air quality data

Particulate matter (PM_{10}) data were provided by the air quality located in Badajoz, which belongs to the surveillance network of Extremadura region (REPICA). It is a suburban unit located at the Badajoz campus of the University of Extremadura (LAT: 38º53’12’’N; LONG: 6º58’15’’O), under the influence of nearby traffic including a highway and urban roads. This study focuses on daily-mean concentration time series of PM_{10} for the period from 1 January 2010 to 31 December 2015. A GRIMM 180

optical laser light scattering aerosol spectrometer (GRIMM Aerosol Technik Ainring GmbH & Co, Germany) was used to record measurements of PM_{10} concentration. The equipment registers 10 minutes averages from which hourly and daily-mean values are calculated, to obtain a database of 2190 day-points (6 years).

The STATISTICA (StatSoft) software was used to analyze the PM_{10} database. The temporal structure of the data was first checked by using the autocorrelation function. Then the overall evolution of the values of the series over time (trend) and its movement in specific periods (seasonality) were explored [6].

3 Results

3.1 Temporal structure

The autocorrelation function shows significant values, $AF_{max} = 0.913$, with a one-day lag. So, it is possible to ensure that the data series has a temporary structure.

3.2 Analysis of Trends

Trend analysis was carried out after applying a smoothing of the experimental data by the moving average method, using a moving average of 3 points. The process was then started to find the mathematical function $T(t)$ which describes the trend of PM_{10} time series. In this case, the model of trend that best fits the time series of PM_{10} in Badajoz is the linear model, following the equation PM_{10} ($\mu g/m^3$) = 19.6 - 0.0025·X (days), ($r = -0.1517$; $p < 10^{-4}$). These results suggest that the decreasing levels of PM_{10} in Badajoz may be attributable to decreased traffic volumes, along with the effects of stricter governmental regulation and improvements to vehicle engine performance, including the fitting of devices for exhaust emission reduction. Steady and even decreasing human activities due to economic crisis in Spain (as in other areas of the European Union) during the studied period may have also contributed to reduced urban PM levels in Badajoz. We'll explore in detail these and additional factors (e.g. meteorological parameters and long-range transport of Saharan dust during the studied period) in future work.

3.3 Analysis of Seasonality

Analysis of seasonal nature of the time series of PM_{10} was carried out by an analysis of the previously obtained (3.1) autocorrelation function and the corresponding periodogram based on a Fourier spectral analysis.

The autocorrelation function shows a cycle that repeats every 355 days (maximum at lag 355) in the range of data. The periodogram significantly confirm the existence of an annual cycle located at a frequency $\omega = 0.0027$ days^{-1}, corresponding to a period of 364.7 days. Considering that a year has 365 days (0.46% error), it is concluded that this time series model predicts the existence of an annual cycle (365 ± 2) days.

3.4 Analysis of outliers

Knowledge trend's function of the time series allows establishing criteria to detect anomalous values of concentration. This information is very useful to know the existence of natural or anthropogenic episodes of pollution which can affect PM concentration levels, even though these peak levels are considered legally admissible.

For this purpose, the trend of the time series is subtracted from the smoothed data, providing a new series with linear behaviour set to zero (avg = 0) and standard deviation $\sigma = 10.39$. A $\pm 2\sigma$ confidence interval was established around the smoothed series data, and values placed above $+ 2\sigma$ or below $- 2\sigma$ were considered anomalous. Several anomalous values of PM_{10} concentration were identified in different periods of the series and these values may correspond to African dust episodes or forest fires that occurred during those years. Work in progress is being done to specify and analyze these values and their origin in detail.

4 Conclusions

Temporal evolution of the concentration of PM_{10} in the ambient air of the city of Badajoz in the period comprising the years 2010 to 2015 was analyzed with statistical techniques of time series.

It is possible to conclude that the evolution of the PM_{10} levels in Badajoz during the studied period presents a structure of time series following a cyclic behaviour with an annual oscillation period every 365 ± 2 day. On the other hand, PM_{10} levels decreased linearly during the period studied at a rate of 0.0025 $\mu g \cdot m^{-3}$ per day.

Anomalous peak values of PM_{10} concentration identified throughout the series, could be related with Saharan dust outbreak episodes or forest fires. A more in-depth study on the behaviour of PM concentrations and causes of variability at different monitoring stations in Extremadura is in progress.

5 Acknowledgements

We acknowledge Junta de Extremadura (projects IB16114 and GR18068), and the Air Quality Surveillance Network of Extremadura (REPICA) , all partially co-financed by the European Funds for Regional Development (FEDER).

References

1. World Health Organization (WHO). Ambient (outdoor) air quality and health. Fact sheet, 313 (2014). http://www.who.int/mediacentre/factsheets/fs313/en/
2. Hara, K.; Homma, J.; Tamura, K.; Inoue, M.; Karita, K.; Yano, E. Decreasing trends of suspended particulate matter and $PM_{2.5}$ concentrations in Tokyo, 1990-2010. Journal of the Air & Waste Management Association, 63, 737-748 (2013).

3. Kassomenos, P.; Vardoulakis, S.; Chaloulakou, A.; Grivas, G.; Borge, R.; Lumbreras, J. Levels, sources and seasonality of coarse particles (PM_{10}–$PM_{2.5}$) in three European capitals – Implications for particulate pollution control. Atmospheric Environment, 54, 337-347 (2012).
4. Harrison, R.M.; Laxen, D.; Moorcroft, S.; Laxen, K. Processes affecting concentrations of fine particulate matter ($PM_{2.5}$) in the UK atmosphere. Atmospheric Environment, 46, 115-124 (2012).
5. Cruz, A; Alves, C.; Gouveia, S.; Scotto, M.G.; Carmo, M.; Wolterbeek, H. A wavelet-based approach applied to suspended particulate matter time series in Portugal. Air Qual Atmos Health, 9, 847–859 (2016).
6. Aguirre Jaime, A. Introducción al tratamiento de series temporales: aplicaciones a las ciencias de la salud. 3nd edn. Díaz de Santos, Cuba (1994).

Long-term (2010-2015) tropospheric ozone temporal series in Badajoz (Spain). Trend and seasonal behavior.

M. Cerrato-Alvarez[1[0000-0002-4057-4493]], C. Miró-Rodríguez[2,* [0000-0002-2270-0871]] and E. Pinilla-Gil[1[0000-0001-5873-7580]]

[1] Department Analytical Chemistry and Research Institute on Water, Climate Change and Sustainability (IACYS), University of Extremadura, Badajoz, 06006, Spain
{macerratoa@unex.es; epinilla@unex.es}
[2] Department of Applied Physics, University of Extremadura, Caceres, 10071, Spain
*cmiro@unex.es

Abstract. Tropospheric ozone is an atmospheric pollutant of great concern, with demonstrated negative effects on human health and the environment. As a secondary pollutant, ozone is produced by complex processes involving reactions of primary emissions with sunlight. In Southern Europe, summer ozone levels are of great concern due to a mix of natural (high temperature and solar irradiation) and anthropogenic factors.
In this work, temporal evolution of ozone levels in Badajoz (Spain) during the period 2010 to 2015 was analyzed. For it, mathematical time-series techniques were used. Firstly, analytic expression for the temporal trend was obtained. This allowed us to estimate that the ozone concentrations decrease with time at a constant rate of -0.0093 (μg m^{-3}) /day.
Lastly, a study of the seasonality of the time-series allowed us to determine that the ozone concentrations varied cyclically. The periods of these oscillations were 24 hours (based on hourly data) and 356 days (based on daily data).

Keywords: tropospheric ozone, time-series, temporal trend, seasonality.

1 Introduction

Tropospheric ozone is an atmospheric secondary pollutant formed by photochemical reactions of volatile compounds and nitrogen oxides activated by solar radiation and temperature [1]. O_3 is produced by the reaction of an oxygen molecule (O_2) with an oxygen atom, which originates from the photolysis of nitrogen dioxide (NO_2) by solar radiation. O_3 is destroyed by reacting with NO to form NO_2 and O_2. While this so-called 'null cycle' does not lead to a net production or destruction of O_3, the presence of volatile organic compounds (VOCs) in the atmosphere interacts with this mechanism through reactions driven by the hydroxyl radial (OH) leading to reactions with the present NO and therefore, to accumulation of O_3 [2].
Periodic episodes of high ozone concentrations are related to negative effects on public health, damage on vegetation and materials degradation. It is also an important

tropospheric greenhouse gas and is referred to as a short-lived climate pollutant [3].

Depending on weather conditions, O_3 precursors can be transported over long distance and originate O_3 formation in locations far from their sources. Several studies highlight the importance of meteorological factors in O_3 formation and transport. These studies have shown O_3 concentrations increase with high temperatures and solar radiation intensities [4]. Other authors have studied daily and annual variations of ozone levels obtaining patterns well defined [5].

Therefore, evolution of local atmospheric conditions combined with variable anthropogenic activities at the emission sources determine production and transport of ozone, resulting in complex, time-dependent concentration profiles. Chemometric and modeling tools are thus essential for proper interpretation of these profiles [6]. The main objective of this work was to analyze the temporal evolution of tropospheric ozone in Badajoz (Spain), using mathematical technique based on time-series.

2 Material and methods

2.1 Data collection

Ozone concentrations were collected from January 2010 to December 2015 at the automatic air quality monitoring station of Badajoz, which belongs to the air quality surveillance network of the regional government of Extremadura. The monitoring station (6º58'15"W; 38º53'12"N) is located at the sports facilities, University of Extremadura, in the northwest suburban area of the city. The measuring point is situated close to traffic sources as a highway and some urban roads. The climate in Badajoz is moderate, between temperate and warm, with low rainfall (523 mm/year). The orographic features and climate play an important role in the interpretation and understanding of ozone behaviour. Ozone measurements were taken by an ultraviolet absorption-based technique, using a commercial UV photometric ozone analyser (Model 49i, Thermo Electron Corporation). This instrument is routinely calibrated and maintained according to standard protocols of European air quality networks. Ozone levels were extracted from the air quality network database as average hourly and daily values.

Calculations of temporal trend and seasonality during the interval 2010 to 2015 were performed using the STATISTICA (StatSoft) software.

3 Results and discussion

3.1 Temporal trend

The trend is the component of the series that describes the direction of the global evolution of the values over time. To study the temporal trend, smoothing of the experimentally daily average data were previously carried out using the moving average method.

The temporal trend of the time series that best fits the behavior of these data is the following linear trend

$$O_3\ (\mu g.m^{-3}) = 63.6691 - 0.0093\ x\ (day) \qquad (r = -0.3106,\ p < 10^{-4})$$

It is concluded that the ozone concentrations decrease with time at a constant rate of -0.0093 (µg m^{-3}) /day.

3.2 Seasonality of the time-series

A study of the seasonal part of the data series has been carried out by analyzing the harmonic components of the set of data. So, we obtained the corresponding periodograms based on a Fourier analysis which assumes that each data time series z (t) is formed by the superposition of sinusoidal components of different frequencies.

Firstly, the periodogram of the spectral content of the hourly data was mapped out, showing a frequency peak $\omega = 0.042\ h^{-1}$ that corresponds to a period of 24 hours. Secondly, the periodogram of the daily data was mapped out, obtaining a frequency peak $\omega = 0.0027\ d^{-1}$, which corresponds to a period of 365 days. These daily and annual patterns for tropospheric ozone have been described in the literature [7].

4 Conclusions

Temporal evolution of ozone levels in Badajoz during the period 2010 to 2015 was analyzed using mathematical technique based on time-series. From these analyses it is possible to conclude that the ozone trend decrease with time with time at a constant rate of -0.0093 (µg m^{-3}) /day and the ozone series is periodic with two different cycles, one daily and other annual probably due to solar radiation the first and the temperature the second.

Acknowledgements

We acknowledge Junta de Extremadura, Spain (projects IB16114 and GR18068), and the air quality surveillance network of Extremadura (REPICA), all partially financed by European Union Funds for Regional Development (FEDER).

References

1. García, G., Allen, A., Cardoso, A.: Development of a sensitive passive sampler using indigotrisulfonate for determination of tropospheric ozone. Journal of Environmental Monitoring 12, 1325–1329 (2010).
2. Im, U., Incecik, S., Guler, M., Tek, A., Topcu, S., Unal, Y., Yenigun, O., Kindap, T., Odman, M.T., Tayanc, M.: Analysis of surface ozone and nitrogen oxides at urban, semi-rural and rural sites in Istanbul, Turkedy. Science of the Total Environment 443, 920-931 (2013).

3. Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E., Stevenson, D.S., Tarasova, O., Thouret, V., Schneidemesser, E., Sommariva R., Wild, O., Williams, M.L.: Tropospheric ozone and its precursors from urban to the global scale from air quality to short-lived climate forcer. Atmospheric Chemistry and Physic 15, 8889-8973 (2015).
4. Agudelo, D.M., Calesso, E., Norte, F.: Time-series analysis of surface ozone and nitrogen oxides concentrations in an urban area at Brazil. Atmospheric Pollution Research 5, 411-420 (2014).
5. Sebald, L., Treffeisen, R., Reimer, E., Hies, T.: Spectral analysis of air pollutants. Part 2: ozone time series. Atmospheric Environment 34, 3503-3509 (2000).
6. Miró Rodríguez, C., Cereceda-Balic, F., Rubio, M.A., Calvo Blazquez, L., Pinilla Gil, E.: Estudio comparativo entre los niveles de ozono troposférico en Santiago (Chile) y Extremadura (España) durante el año 2004. XXVII Jornadas Chilenas de Química. Comunicación QAA 329, Termas de Chillán, Chile (2007).
7. Carnero, J.A.A., Bolívar, J.P., de la Morena, B.A.: Surface ozone measurements in the southwest of the Iberian Peninsula (Huelva, Spain). Environmental Science Pollution Research 17, 355–368 (2010).

Verification on winter rapeseed (*Brassica napus* L.) aboveground dry weight and yield models under waterlogging stress at anthesis

CAO Hong-xin[1], YANG Tai-ming[2,3], ZHANG Bao-jun[4], JIANG Yue-lin[5], GE Dao-ke[1], ZHANG Wen-yu[1], ZHANG Wei-Xin[1], LIU Yan[1], LIU Rui-na[2,3], LI Long[6], WANG Zhou-qin[6], FU Kun-ya[1], FENG Chun-huan[1], CHEN Wei-tao[1] SONG Chu-wei[1], GE Si-jun[1], ZHANG Qian[1], WAN Qian[1]

([1]Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences/Engineering Research Center for Digital Agriculture, Nanjing 210014, China; [2]Institute of Meteorological Sciences of Anhui Province, Hefei 230031, China; [3]Key Laboratory for Atmospheric Sciences & Remote Sensing of Anhui Province, Hefei 230031, China; [4]Northwest Agriculture and Forestry University, Yangling 712100, Shaanxi, China; [5]College of Resources & Environment，Anhui Agricultural University，Hefei，Anhui 230036, China; [6]Experimental station of Xuancheng, Xuancheng 242000, Anhui, China)

Abstract: Waterlogging is a kind of meteorological disasters in rapeseed production in Yangze river valley, in order to quantify the rapeseed growth and yield under waterlogging stress condition, and forecast and control the effect of waterlogging stress on rapeseed, the field, pot, and pool experiments on soil water control were carried out in 2011-2012 taking ZP4 and HYZ16, 2014-2015 and 2015-2016 taking NY18, and NZ19 as materials, respectively. The waterlogging impact factors for the aboveground dry weight per plant, and yield were proposed, and the models for the aboveground dry weight per plant, and yield per plant were initiatively developed through analyzing the effect of waterlogging at anthesis on rapeseed growth and yield for ZP4 and NY18. Of them, the waterlogging impact factors for the aboveground dry weight per plant was the quadratic equation, and yield per plant was the linear equation taking waterlogging duration days as independent variable, respectively. The results showed that the correlation (r), the average absolute difference (d_a), the ratio of d_a to the average observation (d_{ap}), and the RMSE value of simulation and observation in rapeseed aboveground dry weight per plant for NY18, and NZ19 were 0.5731-0.8826, and 0.6538-0.8807 with significant levels at $p<0.05$ or $p<0.01$; 1.2150-2.0142, and 0.0467-1.3383 g plant^{-1}; -5.0731%-0.1462%, and 3.4266%-4.9955%; and 2.2558-3.8426, and 3.0491-4.2795 g plant^{-1} under pot conditions, respectively; 0.6723-0.8768 with significant levels at $p<0.05$ or $p<0.01$, 1.2778-1.6200 g plant^{-1}, 3.2357%-4.2604%, and 2.1498-3.0656 g plant^{-1} under pool conditions, respectively; and that the correlation (r), the average absolute difference (d_a), the ratio of d_a to the average observation (d_{ap}), and the RMSE value of simulation and observation in rapeseed yield per plant for ZP4 under field conditions were 0.7791 with no significant levels at $p<0.05$, 1.0475 g plant^{-1}, -2.5083 g plant^{-1}, -30.0419%, and 4.1948 g plant^{-1}; for NY18, and NZ19 were 0.8500-0.9754, and 0.9555-0.9707 with significant levels at $p<0.01$ levels; -0.31-1.7058, and -0.6308-1.875 g plant^{-1}; -1.4906-5.3551%, and -2.7148%-5.7887; and 1.4827-3.9830, and 1.6217-3.0885 g plant^{-1} under pot conditions, respectively; for NY18, and NZ19 were 0.7674-0.9590 with significant levels at $p<0.01$ levels, -0.2833-1.8478 g plant^{-1}, -1.4513-6.1579%, and 2.1461-3.8905 g plant^{-1} under pool conditions, respectively; It was shown that the simulation precision for the aboveground dry weight per plant, and yield per plant were good under the different sites, ways, and cultivars, and the simulation precision for the yield per plant were different, i.e., pot > pool > field, which indicated that the waterloging stress at anthesis under the field experiment condition was different from that of

under the pot experiment condition. The models developed in this paper can be combined with the rapeseed growth model developed by our group, and used in prediction of rapeseed growth and yield with waterlogging at anthesis after further validating and obtaining the region, and site weather forecast, and model parameters.

Key words: rapeseed; anthesis; waterlogging impact factors; growth and yield; simulation

On the Impact of Shale Oil Revolution in Oil-Dollar Comovement

François Benhmad [1 *]

Abstract:

This paper aims to re-investigate the link between oil prices and US-dollar taking into account the US shale revolution. In order to disentangle their co-movements across different frequency bands, we carry out a wavelet decomposition. Moreover, we also apply Granger causality test to the decomposed time series on a frequency band by frequency band basis during two periods before and after 2008 corresponding to the beginning of oil shale revolution in USA.
We find that at the time scale between 2 and 4 months, there is no Granger causality between US dollar and oil prices. At the seasonal frequency band, there is a bidirectional linear Granger causality between oil prices and US dollar. However, for business cycle frequency, the direction of linear causality have switched throughout the two periods of our dataset. Indeed, from 2000-2007 causality is running from US dollar to oil prices. However, from 2008 to 2017, Granger causality runs from oil prices to US dollar. Hence, shale revolution may have changed the link between oil and the dollar.

Keywords: Oil, US Dollar, Shale Oil, Wavelet Decomposition

JEL codes: E32, Q43
Key Words: shale oil; oil price; US dollar; Granger causality, wavelet decomposition.

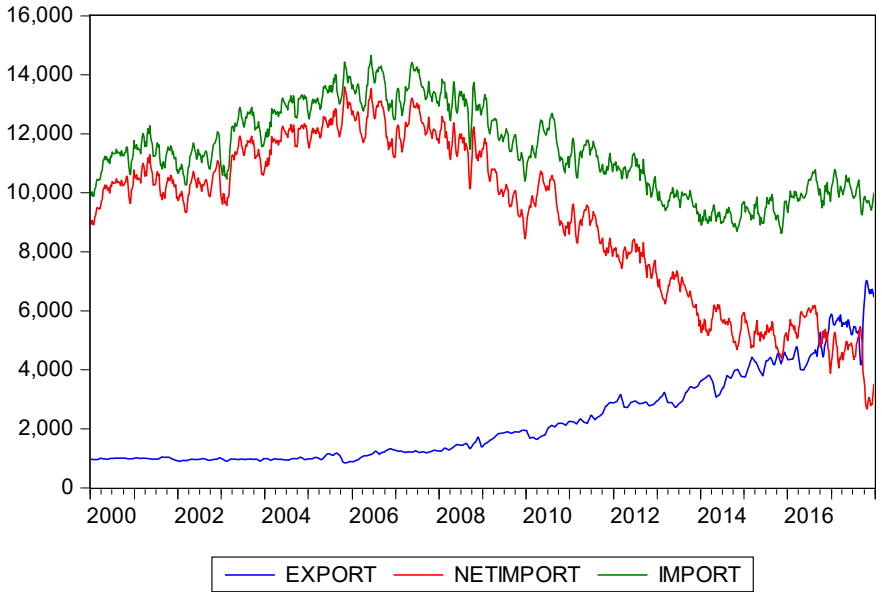
[1 *] Corresponding author, Montpellier University, Site Richter, Avenue Raymond Dugrand, CS79606 , 34960 Montpellier Cedex2, France, Tél: +33.4.34.43.25.02, E-mail address : francois.benhmad@univ-montp1.fr

Extended abstract:

A large empirical literature had explored the comovements between oil prices and US dollar, using mainly the cointegrated VAR models and Granger causality tests. The empirical evidence have produced mixed and inconclusive results. Some authors found a causality running from real oil prices to US Dollar exchange rate (Throop 1993, Zhou 1995, Amano and van Norden 1998a, b, Dibooglu 1995, Chaudhuri and Daniel 1998, Lizardo and Mollick 2010, Chen and Chen 2007, Bénassy-Quéré et al. 2007, Coudert et al. 2008). On the other hand, other studies showed that movements in the U.S. dollar Granger causes crude oil prices (Sadorsky 2000, Yousefi and Wirjanto 2004, Zhang and Wei 2010).

This paper empirically reinvestigates the comovements between oil prices and U.S. dollar taking into account the sharp expansion in US shale oil production that induced a decline in US imports of crude oil (see figure 1).

Figure 1 .U.S. Oil: Imports and Exports (Thousand Barrels per Day)



The goal is to explore the impact of US shale oil on relationship between oil prices and US dollar . Indeed, the shale revolution enables USA to reduce its dependence on imported crude oil whereas it was the world largest importer of crude oil (a share of 20% of world oil imports) before 2008. Saudi Arabia and OPEC oil cartel, which historically had played a swing producer role in oil markets, shifted toward defending market share as they could not agree a cut in oil production. Moreover, the weak demand for oil due to the state of the global economy exacerbates the downward pressure on oil prices and induces the collapse in crude oil prices since mid-2014.

Our contribution to the existing literature is twofold. First, we re-investigate the oil-dollar link before and after the shale oil revolution. We thus subdivide our data sample to 2 time periods, the "great moderation" period from 2000 to 2007 and the "shale revolution" the period from 2008 to 2017. Second, in contrast to most empirical studies which measure the oil-US dollar link on two time scales (short and long terms), we analyze the co-movement of these variables at different times scales by carrying out a wavelet decomposition. Therefore, decomposing oil and US dollar time series on a frequency band by frequency band basis enables us to detect the multiscale Granger causality and to determine at which frequencies one time series causes another. This is particularly important in the case of oil prices and exchange rates since there are different channels through which one may influence the other at varying time scales. In doing so, we can explore the co-movement dynamics between US dollar exchange rate and oil prices on a scale by scale basis (trend, cycle, season, and noise).

Our findings suggest that causal relationship between the US dollar and oil prices is not constant over time scales as they vary across frequency bands. More specifically, we find that at the first frequency band D1 corresponding to a time scale between 2 and 4 months- which corresponds to a class of traders whose investment horizon is very short, i.e. mainly to speculative trading (noise traders) – there is no linear Granger causality. However, at the frequency band corresponding to data seasonal components, there is a bidirectional linear Granger causality between oil prices and US dollar. Then, for time horizon from 16 to 128 months representing approximately the business cycle frequency (fundamentalist traders, especially fund managers and institutional investors), the direction of causality is unstable as we have observed a switching across the two periods of data. Indeed, from 2000-2007 (great moderation) causality is running from US dollar to oil prices, however from 2008 to 2017 (Shale revolution), oil causes US dollar. This switching of causality directions can be linked to shale oil revolution that impacted heavily the oil markets since mid-2014.

Conclusion and Policy Implications

This paper analyzes the co-movements of oil prices and the US dollar by testing linear Granger causality link between their frequency components (high frequency noise, seasonal, business cycle) from 2000 to 2007 and 2008 to 2017. As the two series may be linked through different channels, it is reasonable to expect that they contain signals working at different frequencies. The wavelet decomposition of oil prices and US dollar allows us to provide a refined picture of co-movement analysis as we explored a multiscale causality relationships. From a business cycle perspective, in the sub-sample from 2000_{M1} to 2007_{M12}, the US dollar was the leading variable as the linear causality runs from US dollar to oil prices. However,

after the beginning of shale oil revolution, there was a switch in the causal link, and oil prices become the leading variable. This switching causal relationship is consistent with findings in several other countries (see for example Turhan et al., 2013), but the existing literature falls short of explaining the reasons for this switch. We argue that shale oil revolution provides a reasonable explanation.

Indeed, shale oil revolution is a major change in international oil markets as the sharp expansion in US shale oil production enables USA to reduce its dependence on imported crude oil. Moreover, before 2008, USA was the world largest importer of crude oil with a share of 20% of world oil imports. As the oil is invoiced predominantly in US dollars, the shale oil boom not only impacted the oil prices but also its relationship with the US dollar as was shown by the causality direction switch after the beginning of shale oil revolution making oil prices leading variable.

The results reported in this paper can be extended in several ways in future research. For instance, given the rapid increase of emergent economies share of the overall energy demand and the world economic growth, it would be interesting to investigate the cyclical co-movements and causality relationship between world demand of oil, oil prices and U.S dollar in order to deepen our understanding to the link between oil prices and U.S dollar. Another interesting topic for future research, using the same methodology, the dynamic cyclical co-movements of oil prices can be empirically investigated with other macroeconomic variables such as consumer prices, unemployment, and stock prices, implied stock markets volatility (VIX), monetary policy variables (federal funds) as well as with other commodity prices.

References

Ahamada, I. and Jolivaldt, P. (2010). Classical vs wavelet-based filters Comparative study and application to business cycle. *Document de Travail du Centre d'Economie de la Sorbonne - 2010.27.*

Almasri, A. and Shukur, G. (2003). An Illustration of the Causality Relationship between Government Spending and Revenue Using Wavelets Analysis on Finnish Data. *Journal of Applied Statistics*, 30, 571-584.

Aloui, R., Aïssa, M. S. B. and Nguyen, D. K. (2013). Conditional dependence structure between oil prices and exchange rates: A copula-GARCH approach. Journal of International Money and Finance 32, 719-738.

Amano, R. A. and van Norden, S. (1998a). Oil price and the rise and fall of the U.S. real exchange rate. *Journal of International Money and Finance*, 17, 299-316.

Amano, R. A. and van Norden, S. (1998b). Exchange Rates and Oil price. *Review of International Economics*, 6, 683-94.

Babajide, F. (2014). Modelling the oil price–exchange rate nexus for South Africa. International Economics, 140, 36-48.

Bal, P. D. and Badri, N. R. (2015). Nonlinear causality between crude oil price and exchange rate: A comparative study of China and India. Energy Economics 51, 149-156.
Baumeister, C. and Kilian, L. (2016). Lower Oil Prices and the U.S. Economy: Is This Time Different? http://ceepr.mit.edu/files/papers/2016-014.pdf
Baxter, M. (1994). Real exchange rates and real interest differentials Have we missed the business-cycle relationship? *Journal of Monetary Economics 33, 5-37.*
Bénassy-Quéré, A., Mignon, V. and Penot, A. (2007). China and the relationship between the Oil price and the Dollar. *Energy Policy*, 35, 5795-5805.
Benhmad, F. (2012). Modelling the nonlinear causality test between oil price and US dollar: a wavelet approach. *Economic Modeling*, 29, 1505-1514.
Benhmad, F. (2013). Dynamic cyclical comovements between oil prices and US GDP: a wavelet perspective. Energy Policy, 57, 141-151.
Bernanke, A. (2004). Oil and the Economy. Speech presented at Darton College, Albany, GA, *http://www.federalreserve.gov/boarddocs/speeches/2004/20041021/default.htm*.
Blanchard, O. J. and Gali, J. (2007). The Macroeconomic Effects of Oil Price Shocks: Why Are the 2000s So Different from the 1970s? *Working paper 07-21. Cambridge, MA: Massachusetts Institute of Technology, Department of Economics.*
Bouoiyour, J., Selmi, R., Tiwari, A. K. and Shahbaz, M. (2015). The nexus between oil price and Russia's real exchange rate: Better paths via unconditional vs conditional analysis. *Energy Economics* 51, 54-66.
Brahmasrene, T., Huang, J-C. and Yaya, S. (2014). Crude oil prices and exchange rates: Causality, variance decomposition and impulse response. Energy Economics 44, 407-412.
Burns and Mitchell (1946). Measuring Business-Cycles". New York: *National Bureau of Economic Research*
Chaudhuri, K. and Daniel, B. C. (1998). Long run equilibrium real exchange rates and oil price. Economics Letters, *58*, 231-238.
Chen, H., Liu, L., Wang, Y. and Zhu, Y. (2016). Oil price shocks and U.S. dollar exchange rates. Energy 112, 1036-1048.
Chen, S. S. and Chen, H. C. (2007). Oil price and real exchange rates. *Energy Economics*, 29, 390-404.
Coudert, V. and Mignon, V. (2016). Reassessing the empirical relationship between the oil price and the dollar. Energy Policy, 95, 147-157.
Coudert, V., Mignon, V. and Penot, A. (2008). Oil price and the Dollar. *Energy Studies Review*, 18, 2.
Crowley, P. (2007). A guide to wavelets for economists. Journal of Economic Surveys, 21, 207-267.
Crowley, P. (2010). Long cycles in growth: explorations using new frequency domain techniques with US data. Bank of Finland Research Discussion Papers, 6/2010.
Crowley, P. and Lee, J. (2005). Decomposing the co-movement of the business cycle: A time-frequency analysis of growth cycles in the euro area. Bank of Finland Discussion Paper 12/2005, Helsinki, Finland.
Dalkir, M. (2004). A new approach to causality in the frequency domain. *Economics Bulletin*, 44, 1-14.
Daubechies, I. (1992). *Ten lectures on wavelets.* CBMS-NSF conference series in applied mathematics, SIAM Ed.
De Truchis, G. and Keddad, B. (2016). On the risk comovements between the crude oil market and U.S. dollar exchange rates. Economic Modelling 52, 206-215.
Dhamala, M., Govindan, R. and Mingzhou, D. (2008). Estimating Granger causality from Fourier and wavelet transforms of time series data. *Physical Review Letters*, 100, 018701

Dibooglu, S. (1995). Real disturbances, relative prices, and the purchasing power parity. *Journal of Macroeconomics, 18, 69-87.*
Economic Dynamics and Control, 2, 329-352.
Ewing, B. and Thompson, M. A. (2007). Dynamic cyclical comovements of oil prices with industrial production, consumer prices, unemployment, and stock prices. *Energy Policy, 35, 5535-5540.*
Fratzscher, M., Schneider, D. and Van Robays, I. (2014). Oil prices, exchange rates and asset prices. European Central Bank, *Working Paper* N°1689, July.
Gallegati, M. and Gallegati, M. (2007). Wavelet variance analysis of output in G7 countries.
Gençay et al. (2002)''Multiscale systematic risk, National center of competence in research financial valuation and risk management'', *Working Paper n°43*
Gençay, R., Gradojevic, N., Selcuk, F. and Whitcher, B. (2010). Asymmetry of information flow between volatilities across time scales. Quantitative Finance, forthcoming.
Gillman, M. and Nakov, A. (2009). Monetary effects on nominal oil price. *The North American Journal of Economics and Finance*, 20, 239-254.
Golub, S. S. (1983). Oil price and exchange rates. *The Economic Journal*, 371, 576-593.
Granger, C. W. J. (1969). Investigating causal relation by econometric and cross-sectional method. *Econometrica,* 37, 424-38.
Granger, C. W. J. (1980). Testing for causality: a personal viewpoint. *Journal of*
Grossman. A. and Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15, 723-736.
Habib, M. M., Butzer, S. and Stracca, L. (2016). Global exchange rate configurations: Do oil shocks matter? IMF Economic Review 64, 443-470.
Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, 113, 363-398.
Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007-08. *Brookings Papers on Economic Activity*, 1, 215-261.
Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91, 228-248.
Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics*, 38, 215-220.
Hamilton, J. D. (2008). Oil and the macroeconomy, In: S. Durlauf and L. Blume (Eds.), *New Palgrave Dictionary of Economics and the Law, 2nd ed., McMillan.*
Hamilton, J. D. (2010). Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic Dynamics*, 15, 364-378.
Hasanov, F., Mikayilov, J., Bulut, C., Suleymanov, E. and Aliyev, F. (2017). The role of oil prices in exchange rate movements: The CIS oil exporters. Economies 13, 1-18.
In, F. and Kim, S. (2006). The hedge ratio and the empirical relationship between the stock and futures markets: A new approach using wavelets. *The Journal of Business*, 79, 799-820.
Jammazi, R., Lahiani, A. and Nguyen, D. K. (2014). A wavelet-based nonlinear ARDL model for assessing the exchange rate pass-through to crude oil prices. Journal of International Financial Markets, Institutions and Money 34, 173-187.
Jawadi, F., Louhichi, W., Ameur, H. B. and Cheffou, A. I. (2016). On oil-US exchange rate volatility relationships: An intraday analysis. Economic Modelling 59, 329-334.
Kilian, L. (2016). The impact of the Shale oil revolution on U.S. oil and gasoline prices. Review of Environmental Economics and Policy 10, 185-205.
Kim, J-M. and Jung, H. (2018). Relationship between oil price and exchange rate by FDA and copula. Applied Economics 50, 2486-2499.
Kim, S. and In, H. F. (2003). The relationship between financial variables and real economic activity: evidence from spectral and wavelet analyses. *Studies in Nonlinear Dynamics and Econometrics*, 7, 1-18.

Krugman, P. (1983a). Oil and the dollar, In: Bhandari, J.S., Putman, B.H. (Eds), *Economic Interdependence and Flexible Exchange rates*, MIT Press, Cambridge, MA.
Krugman, P. (1983b). Oil shocks and exchange rate dynamics, In: Frenkel, J.A. (Ed.), *Exchange Rates and International Macroeconomics*, University of Chicago Press, Chicago.
Lee, T. H., White, H. and Granger, C. W. J. (1993). Testing for neglecting nonlinearity in time series models. *Journal of Econometrics*, 56, 269-290.
Lescaroux, F. and V Mignon, V. (2008). On the influence of oil prices on economic activity and other macroeconomic and financial variables. OPEC Energy Review, 32, 343-380.
Li, J. (2006). Testing Granger Causality in the presence of threshold effects. *International Journal of Forecasting*, 22, 771-780.
Li. X-P., Zhou, C-Y. and Wu, C-F. (2017). Jump spillover between oil prices and exchange rates. Physica A 486, 656-667.
Lizardo, R. A. and Mollick, A. V. (2010). Oil price fluctuations and U.S. dollar exchange rates. *Energy Economics*, 32, 399-408
Mallat, S. (1989). A theory for multiresolution signal decomposition, the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674-93.
Manescu, C. B. and Nuno, G. (2015). Quantitative effects of the shale oil revolution. *Energy Policy*, 86, 855-866.
McLeod, R. C. D. and Haughton, A. Y. (2018). The value of the US Dollar and its impact on oil prices: Evidence from a non-linear asymmetric cointegration approach. Energy Economics 70, 61-69.
Mensi, W., Hammoudeh, S. and Seong-Min, Y. (2014). Structural breaks, dynamic correlations, asymmetric volatility transmission, and hedging strategies for petroleum prices and USD exchange rate. Energy Economics 48, 46-60.
Mensi, W., Hammoudeh, S., Shahzad, S. J. H., Al-Yahyaee, K. H. and Shahbaz, M. (2017). Oil and foreign exchange market tail dependence and risk spillovers for MENA, emerging and developed countries: VMD decomposition based copulas. *Energy Economics 67*, 476-495.
Mitra, S. (2006). A wavelet filtering based analysis of macroeconomic indicators: the Indian evidence. *Applied Mathematics and Computation*, 175, 1055-1079.
Mork, K. A. (1989). Oil and the macroeconomy when prices go up and down: An extension of Hamilton's results. *Journal of Political Economy*, 91, 740-744.
Nikanor, I. V. and Yuhn, K-Y. (2016). Oil price shocks and exchange rate movements. Global Finance Journal 31, 18-30.
Percival, D. B. and Walden, A. T. (2000). *Wavelet methods for time series analysis*. Cambridge University Press, 2000.
Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society*, 27, 204-237.
Priestley, M. B. (1996). Wavelets and time-dependent spectral analysis. *Journal of Time Series Analysis*, 17, 85-103.
Ramsey, J. B. and Lampart, C. (1998). Decomposition of economic relationships by timescale using wavelets. *Macroeconomic Dynamics*, 2, 49-71.
Reboredo, J. C. (2012). Modelling oil price and exchange rate co-movements. Journal of Policy Modeling, 34, 419-440.
Reboredo, J. C. and Rivera-Castro, M. A. (2013). A wavelet decomposition approach to crude oil price and exchange rate dependence. Economic Modelling 32, 42-57.
Rogoff, K. (1991). Oil productivity, government spending and the real yen-dollar exchange rate. Federal Reserve Bank of San Francisco Working Papers 91-06.
Sadorsky, P. (2000). The empirical relationship between energy futures prices and exchange rates. *Energy Economics*, 22, 253-266.

Schleicher, C. (2002). An Introduction to Wavelets for Economists. Working Papers 02-3, Bank of Canada.
Studies in Nonlinear Dynamics Econometrics, 11, 1435-1455.
Throop, A. (1993). A generalized uncovered interest parity model of exchange rates. *Federal Reserve Bank of San Francisco Economic Review*, 2, 3-16.
Tiwari, A. K. and Albulescu, C. T. (2016). Oil price and exchange rate in India: Fresh evidence from continuous wavelet approach and asymmetric, multi-horizon Granger-causality tests. Applied Energy 179, 272-283.
Turhan, I., Hacihasanoglu, E. and Soytas, U. (2013). Oil prices and emerging market exchange rates. Emerging Markets Finance and Trade 49, 21-36.
Turhan, M. I., Sensoy, A., Hacihasanoglu, E. (2014). A comparative analysis of the dynamic relationship between oil prices and exchange rates. Journal of International Financial Markets, Institutions and Money 32, 397-414.
Uddin, G. S., Tiwari, A. K., Arouri, M. and Teulon, F. (2013). On the relationship between oil price and exchange rates: A wavelet analysis**.** Economic Modelling 35, 502-507.
Wen, F., Xiao, J., Huang, C. and Xia, X. (2017). Interaction between oil and US dollar exchange rate: nonlinear causality, time-varying influence and structural breaks in volatility. Applied Economics, 319-334.
Yogo, M. (2008). Measuring Business Cycles: a wavelet Analysis of Economic Time Series. *Economics Letters*, 6, 208-212.
Yousefi, A. and Wirjanto, T. S. (2004). The empirical role of the exchange rate on the crude-oil price forward. *Energy Economics*, 26, 783-799.
Zhang, C. and Farley, A. (2004). A Multiscaling Test of Causality Effects among International Stock Markets. *Neural, Parellel and Scientific Computations*, 12, 91-112.
Zhang, Y. J. and Wei, Y. M. (2010). The crude oil market and the gold market: evidence for cointegration, causality and price discovery. *Resources Policy*, 35, 168-177.
Zhang, Y. J., Fan, Y., Tsai, H. T. and Wei, Y. M. (2008). Spillover effect of U.S. dollar exchange rate on oil price. *Journal of Policy Modeling*, 30, 973-991.
Zhang, Y-J., Fan, Y., Tsai, H-T. and Wei, Y-M. (2008). Spillover effect of US dollar exchange rate on oil prices. Journal of Policy Modeling 30, 973-991.
Zhou, S. (1995). The response of real exchange rates to various economic shocks. *Southern Economic Journal*, 61, 936-954.

Forecasting inflation with long-short term memory recurrent neural networks: the Colombian case

Andres C. Serna, Javier G. Diaz, and Julio C. Alonso

Universidad Icesi, Cali, Colombia
acserna@icesi.edu.co
jgdiaz@icesi.edu.co
jcalonso@icesi.edu.co

1 Abstract

With the latest developments of deep learning models, new approaches on recurrent neural networks have pointed to new ways to improve forecasting accuracy. More specifically, Long-Short Term Memory (LSTM) models provide an alternative method to analyze economic time series data.

Artificial or recurrent neural network methods have been applied to forecasting, analyzing macroeconomic time series. In this paper, we examine the accuracy of LSTMs, a type of recurrent neural networks (RNN), and more traditional artificial neural networks (ANN) to forecast Colombias inflation rate and we propose a methodology to select the best models to estimate Colombias inflation.

We compare the performance of several ensembles of neural networks models, with different architectures with the traditional methods. The traditional methods considered are exponential smoothing (ETS), autoregressive integrated moving average (ARIMA), Holt-Winters, Prophet and a hierarchical time series approach for disaggregated Consumer Price Index (CPI. We compare the models using the RMSE metric and the smallest error convergence (using segmented regression and statistical tests).

Inflation is one of the most fundamental variables in macroeconomics. Both governments and private sector require reliable inflation predictions to make accurate and useful decisions and take appropriate measures. Inflation forecasting is necessary to assess business opportunities, design long-term growth strategies, anticipate possible financial risks, among other tasks in the business sector. To central banks, the institutions responsible for managing the monetary policy regime of inflation targeting (Colombias case), inflation forecasting is also valuable because it is one of the factors driving monetary policy. Indeed, it is important to keep inflation low and stable, since a high inflation has undesirable effects on the economy; for example, it harms the purchasing power of the

national currency, it affects economic growth (as investment projects become riskier), distorts consumption and savings decisions, causes an unequal distribution of income and hinders financial intervention.

Our approach implemented ensembles of LSTMs and ANN models, using monthly data of the inflation rate, and the aggregated CPI index and disaggregated CPI index by the household level of income and city. Our methodology consists of the following steps. First, we select the neural networks that best predict inflation rate by varying a set of hyper-parameters and configurations controlling its architecture for m months and n executions (epochs, in neural networks terms). The second step implies to obtain the validation errors and carry out a segmented regression to choose the most predictive neural networks configurations. Then, the third step implies using the selected configurations to forecast inflation rate, those models with the lowest RMSE conform ensembles of models, whose aggregated forecast is computed by applying linear regression, ridge regression, lasso regression, random forest, gradient boosting and extreme gradient boosting. In the fourth step, we compare our neural network ensemble models forecast with results from several econometric methods and selecting the best models. Finally, we select and validate the best models to forecast the inflation by implementing Diebold-Mariano (1995) statistical test.

The results show that LSTM and ANN models outperform more traditional econometric methods (both using data of inflation rate and disaggregated CPI) in Colombias inflation one step ahead forecast, specifically the LSTM and ANN using data of inflation rate with the best model (minimum RMSE) and the ridge regression ensemble. Also, using LSTM models in the context of Colombian inflation forecasting does not considerably improve predictive accuracy over traditional ANN. Nevertheless, LSTM can still be an attractive alternative, since they can take into account other types of information not considered in our study (such as macroeconomic and weather variables).

Hybrid forecasting methods applied to the Earth's rotation and Radon time-series for anomalies detection

F. Ambrosino[1][0000-0003-2122-9130], L. Thinová[2], M. Briestenský[3], C. Sabbarese[1]

[1] Department of Mathematics and Physics, University of Campania "Luigi Vanvitelli", Viale Lincoln 5, 81100, Caserta, Italy;
fabrizio.ambrosino@unicampania.it
[2] Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 11519, Prague 1, Czech Republic;
[3] Institute of Rock Structure and Mechanics, Czech Academy of Sciences, V Holešovičkách 41, 18209, Prague 8, Czech Republic.

Abstract. This contribution presents three new forecasting hybrid methods for identifying anomalies in time-series, tested in two interesting applications.

An increasing number of signal processing methods have been presented in literature for different purposes (Tascikaraoglu and Uzunoglu, 2014). The major used approaches are: statistics (e.g. ARIMA-auto-regressive integrated moving average, MLR-multiple linear regression); artificial intelligent (ANN-neural networks); spectral decomposition (e.g. SSA-singular spectrum analysis, EMD-empirical mode decomposition); machine learning (e.g. SVR-support vector regression). Recently, hybrid methods are increasingly used because they are made up of two or more traditional individual algorithms, combining the advantages of the single to better extract and capture the characteristics of time-series. These methods have higher accuracy, with an absolute error lower than individual ones (Du et al., 2018).

In the present work, three cutting-edge hybrid models are proposed, developed by homemade algorithm codes in Matlab® computing environment. The combined methods are briefly described as follows.

EMD+SVR. EMD is an adaptive and non-parametric technique that decomposes the signal into several components which represent a certain physical phenomenon occurring. Each component is a function that has only one extreme between zero crossings, along with a mean value of zero (Yu et al., 2017). SVR is another non-parametric algorithm (Sousa et al., 2014) which trains a regression model on each component coming form previous method, in order to predict a function having smallest deviation from the original time-series, as flat as possible. The final result is given by prediction method applied to the combination of the outputs from all sub-series.

MLR+ARIMA. MLR creates a multiple linear regression, performed by least-squares fit, appropriate for studying the dependence relationship between a time-series and the joint distribution of one or more influencing predictor (Sabbarese et al., 2017). The ARIMA method (Yaacob et al., 2010), applied on the residue, captures and forecasts the dependency of the future values on the previous data, according to the autoregressives (AR) and moving average (MA) algorithms by differentiating the time-series (I).

SSA+FM. SSA is a non-parametric spectral estimation technique, which uses a set of input vectors of the time-series and of delayed phase-copies of itself. It consists in the decomposition and the reconstruction of the signal, combining elements of multivariate statistics and multivariate geometry. Equal pairs of eigenvalue are calculated in spectrum of the autocovariance matrix when persistent oscillations are present in the signal (Briceño et al., 2013). The forecasting part (FM), is formulated by modelling the last components of the corrispondent eigenvector (Hassani and Thomakos, 2010).

These methods are applied with the purpose of anomaly detection. A value is considered anomalous when is not within the 95% confidence interval from the original signal; it is positive or negative above or below the upper and lower levels, respectively. This interval level reflects the reliability of results according to proposed forecasting methods. The implemented and appropriately tested methods are applied to two different time-series for the evaluation of the performances and of the lower mean percentage prediction error, in order to select the best hybrid method. The two applications are relative to the variation of Earth's rotation rate and the gas Radon measurement in soil gas, both in connection with Earth's movements (Sabbarese et al., 2017; Wang et al., 2011). The results obtained from the first time-serie prove powerful coincidences with World's daily earthquakes stand for magnitude ≥7 and number of events ≥150. The results of the Radon time-series, recorded in 7 caves of 4 European Countries (Slovakia, Czech Republic, Slovenia, Italy), demonstrate as the gas could be a geochemical precursor of faul displacement, volcanic activity and earthquake in surrounding area. The overall results are supported by statistical analysis, according to cross-correlation factors. Good equivalency is found among the three hybrid methods. EMD+SVR is resulted the best one thanks to the few necessary parameters for the analysis process that make it more stable, whereas MLR+ARIMA exhibits the lower performance because of the opposite reason.

The results from the two applications clearly demonstrate the effectiveness of the applied methods in the finding of anomalies in time-series, and give improvements and innovation in the concerning research. The capability of these methods can be exploited to any kind of time-series of different research sector where it is useful to highlight anomalies in a signal.

Keywords: Hybrid forecasting method, Time-series analysis, Anomalous signal

References

1. Briceño, H., Rocco, C.M., Zio, E., 2013. Singular spectrum analysis for forecasting of electric load demand. Proceedings of PHM Conference, 33, 919–924.
2. Du, P., Wang, J., Yang, W., Niu, T. Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. Renew. Energ. 122, 533-550 (2018).
3. Hassani, H., Thomakos, D. A review on singular spectrum analysis for economic and financial time series. Stat. and its Interface 3, 377-397 (2010).
4. Sabbarese, C., Ambrosino, F., De Cicco, F., Pugliese, M., Quarto, M., Roca, V. Signal decomposition and analysis for the identification of periodic and anomalous phenomena in Radon time-series. Radiat. Prot. Dosim., 177(1-2), 202–206 (2017).

5. Sousa, J.C., Jorge, H.M., Neves, L.P., 2014. Short-term load forecasting based on support vector regression and load profiling. Int. J. Energ. Res. 38(3), 350–362.
6. Tascikaraoglu, A., Uzunoglu, M. A review of combined approaches for prediction of short-term wind speed and power. Renew. Sustain. Energy. Rev. 34, 243–254 (2014).
7. Wang, H., Zhao, X., Li, Y., Chen, X. Relationship between Earth's rotation and several strong earthquakes and moderate-small earthquakes occurring around the epicenter regions pior to strong earthquakes. Earthq. Res. China 25 (4) 443-455 (2011).
8. Yaacob, A., Tan, I. K. T., Chien, S. F., Tan, H. K., 2010. Arima based network anomaly detection. Proceedings of ICCSN Conference 205-209.
9. Yu, C., Li, Y., Zhang, M. Comparative study on three new hybrid models using Elman Neural Network and Empirical Mode Decomposition based technologies improved by Singular Spectrum Analysis for hour-ahead wind speed forecasting. Energy Conv. Manag. 147, 75–85 (2017).

ANALYSES OF THE TIME SERIES BASED ON ATMOSPHERIC ENERGY BUDGET DETERMINATION FOR THE PURPOSE OF BUDGET PROGNOSIS WITH ARMA METHOD

Monika Birylo, University of Warmia and Mazury in Olsztyn, Poland

E-mail: monika.sienkiewicz@uwm.edu.pl

Abstract. The variability of the atmospheric energy budget time series is presented in the paper. The atmospheric energy budget is usually computed as a difference between precipitation (P) and evapotranspiration (ET) over an area. In the research, the area of Poland was considered. Three types of models are tested: the Global Land Assimilation System (GLDAS), The Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2), and Standardized Precipitation Evapotranspiration Index (SPEI). The atmospheric energy budget and TWS (surface plus ground) changes were computed for the period 01.2006 to 12.2017. Based on the computation a prognosis using ARMA (Auto Regressive Integrated Moving Average) method was introduced. After the seasonal decomposition, the following parameters shows the best equivalency of ARMA model and the data: p=2, d=0, q=2.
Based on the research, it can be concluded that the time series comparisons show a good agreement between the GLDAS, MERRA-2 and SPEI models. A comparison between the real time series and the prediction for the ten months in the period of 01.2006-12.2015 proved the assumption that the ARMA model can be a useful tool for time series forecasting.

Keywords: atmospheric energy budget, precipitation, evapotranspiration, time series, prognosis

1 Introduction

A good representation of the earth's climate system are values of an atmospheric energy budget. It can be calculated as a difference between water inflow and water outflow, so, between precipitation and evapotranspiration over a region (AEB=P-ET). Evaluating atmospheric energy budget let us monitor the interaction between land and atmosphere [1]. The data for a computation can be acquired exactly from assimilated models, satellite observations, and the combination of in-situ and satellite measurements. An intensification of the global hydrological cycle is caused by a climate change [2, 3].

The whole area of Poland was analyzed: SW (49.822711°N; 14.996841°E) – NE (54.620711°N; 23.649940°E). The paper is a trial of presenting the surface fluxes of

adfa, p. 1, 2011.

the total water between the surface and the atmosphere and predicting future values of the atmospheric energy budget.

A basis part of the climate system is a stable atmospheric energy budget. As a global warm is a well-known phenomenon, continuous monitoring of the energy budget is very important [4]. The climate change causes plenty of uncertainties of rate of seasonal and annual precipitation and evapotranspiration, which influence atmospheric energy budget determination [5]. That is the reason, why monitoring atmospheric energy budget is under interest of hydrologist and meteorologists over the world. It is especially important in the arid and semi-arid regions where evapotranspiration is over 90% of precipitation, which means a significant loss of a drinking freshwater. Globally, it is calculated that over 60% of precipitation fallen on terrestrial surface returns to atmosphere in a form of evapotranspiration [6].

The principal organization concerning knowledge about time series of the atmospheric energy budget was The World Meteorological Organization's Global Energy and Water Cycle Experiment Continental-Scale International Project (GCIP). The mail objective of the project was developing a complex land–atmosphere–ocean interactions in the basin of Mississippi River. It was concluded that understanding characteristics of the residuals make us possible to prepare critical element of statistical analysis. This assumption makes it possible to identified model deficiencies and biases [7]. In [5] the atmospheric energy budget was compared when calculated over land and over ocean. It was revealed during the research that much larger changes can be noticed over ocean than over land. It can be concluded that ocean energy budget changes follow a thermodynamic scaling of the atmospheric moisture convergence, which is called: "wet-get-wetter, dry-get-drier" mechanism. Over land simple scaling did not find a region with decreasing trend of the atmospheric energy budget. Swenson and Wahr [8] paid attention to the fact that calculating evapotranspiration or potential evapotranspiration, which is unmeasured in-situ makes the budget computations difficult. In the paper [8] continental changes in a large scale of the total water storage form the Gravity Recovery and Climate Experiment (GRACE) are integrated with P-ET.

The statistical analyses of mass variations (presented in a form of the atmospheric energy budget, for example) have to be based on a time series assessment. Here, time need to be independent variable. Such stochastic analysis can give us an opportunity of making some simulations which are based on the tested phenomenon [9]. All stochastic processes can be presented in a form of the time series. Such presentation can give an answer what is the reason of initiating researched process. Unit of the time series determine analysis method. Every stochastic process, every time series consists of:

- trend,
- periodical fluctuations,
- random fluctuations,
- random changes of the process caused by some external influences [10].

2 Methodology

Atmospheric energy budget is the difference between precipitation and evapotranspiration:

AEB=P-ET,

where P is a precipitation, ET – evapotranspiration (or potential evapotranspiration PET).

Evapotranspiration constitutes a significant part of precipitation over the land surface [11].

Prognosis usually are based on present or past events, so are based on a time observations or random variables. If in the set of data, a systematical component is observed, a variable coefficient has to be computed. The coefficient greater than 0.05 means that smoothing need to be introduced (by computing a weighted moving average (WMA)). When every systematic element is inherent (systematical component, like random noise, trend, seasonal fluctuations), a time series decomposition is commonly used. The most effective method in such example is ARIMA/ARMA (Auto Regressive Integrated Moving Average) method [10, 12]. The method consists of three elements:

- Autoregressive process (AR),
- Moving Average process (MA),
- Integration Level [13].

3 Data

The For the atmospheric energy budget computation, we need: precipitation and evapotranspiration. Water in a form of liquid, gas or snow can be in or on land in different forms, stays on a soil, can be infiltrated inside a ground, some of it run-off depending on the basins around, some is accumulated in the plant's leaves and in basins of course. Precipitation is dependent on:

- climate,
- soil texture,
- soil structure,
- depth of the rootzone,
- topography,
- initial soil moisture content,
- and irrigation methods [14].

The sum of evaporation (water leaving the ground) and transpiration (water leaving plants) is the evapotranspiration. Evapotranspiration defined in [15], is a complex empirical function of air temperature and day length. Evapotranspiration depends on:

- climate,
- type of crop,
- and soil moisture [14].

For the purpose of the research three models were used: GLDAS (the Global Land Assimilation System), MERRA-2 (The Modern-Era Retrospective analysis for Re-

search and Applications version 2) and SPEI (Standardized Precipitation Evapotranspiration Index). First two models are presented in a form of fluxes that can be used for the atmospheric energy budget computation, the third model gives us ready time series.

GLDAS model (the Global Land Assimilation System) is worldwide gridded distributed. The model consists of 28 elements which are geophysical parameters. All parameters are organized in a form of four sub-models: VIC (Variable Infiltration Capacity Model), NOAH (National Centres for Environmental Prediction/Oregon State University/Air Force/Hydrologic Research Lab Model), Mosaic Land Surface Model and CLM (Common Land Surface model). GLDAS aim is presenting land surface states and fluxes, it integrates satellite and in-situ measurements. For this purpose, advanced modelling of a land surface is introduced, and data assimilation techniques, as well. The GLDAS model is based on two components: observation-based forcing datasets and meteorological forcing data [16, 17]. GLDAS data was provided by: [18, 19].

A reanalysis model, based on the historical climate analyses is The Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2). The model describes weather and climate conditions in different scales and locations [20]. Data for the MERRA-2 model is obtained form in situ observations, satellite radiances and several remotely sensed data over the era of Earth-observing satellite. The model consists of data needful for water and energy estimation budget in 1-hour range in a regular grid 0.5°×0.67° grid. It need to be noticed that also the influence of stratospheric ozone, land surface improvement, cryospheric processes are considered in the model [21, 22].

Data essential for monitoring draught conditions, in near-real-time is distributed in a form of the Standardized Precipitation Evapotranspiration Index (SPEI). The model is published in the 0,5 degrees spatial resolution and one-month time resolution, covering all the globe for the period 1950-2010, and time scales 1-48 months. The SPEI model consists of precipitation (P) and potential evapotranspiration (PET). The model characterizes combining multi-timescales aspects of the Standardized Precipitation Index (SPI). It is also added evapotranspiration information. These two parameters make the model very useful for climate studies. Temperature data is obtained from the NOAA NCEP CPC GHCN_CAMS gridded dataset, precipitation from the Global Precipitation Climatology Centre (GPCC) [23].

4 Results and discussion

4.1. Atmospheric energy budget determination

For the purpose of the atmospheric energy budget estimation two models were taken: GLDAS, sub-model NOAH and MERRA-2 model. In the research, a time series form a period 01.2006-12.2017 are considered. An upward trend of precipitation and evapotranspiration time series in Poland can be observed in a both models. The high-

est values are noticed in every Julies, near zero at beginning of the springs and the autumns. Based on both models the atmospheric energy budget was computed. I could notice values between -21.08 e-6 kg/m2/s (in 2016.03) and 35.52 e-6 kg/m2/s (in 2010.05) for GLDAS model, and between -14.20 e-6 kg/m2/s (in 2014.07) and 41.35 e-6 kg/m2/s (in 2017.03) for MERRA-2 model in the researched period. Computed atmospheric energy budget was then compared with the SPEI model based on the potential evapotranspiration. In the [24] SPEI1, SPEI 6 and SPEI 12 were compared, giving the most correlated values for SPEI 1. Here, only SPEI 1 model is taken into account. SPEI model gives us ready for use atmospheric energy budget. The values are between -10.95e-6 kg/m2/s (in 2009.04) and 28.33-6 kg/m2/s (in 2010.05) for SPEI1 model. The time series of the computed atmospheric energy budget are presented in the figure 1.

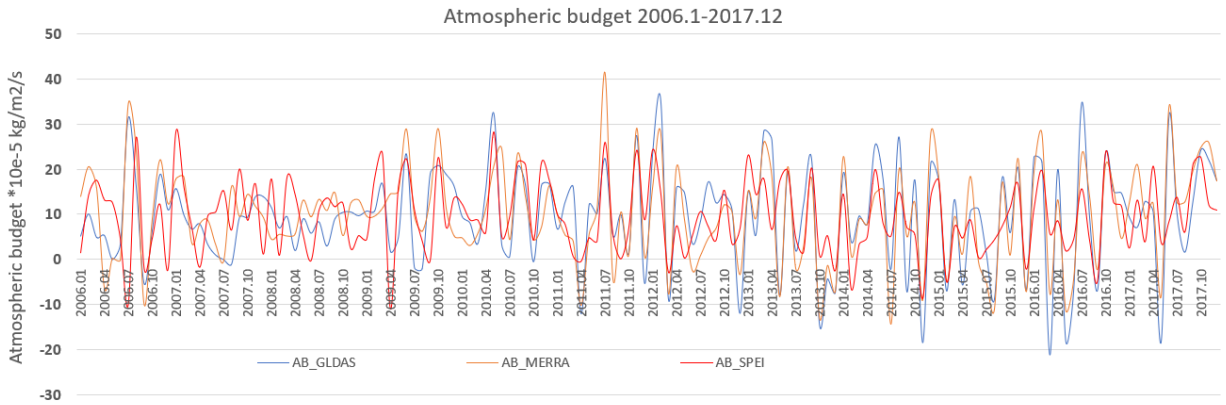


Fig. 1. Atmospheric budget values for the period 2006.01 – 2016.12 computed with GLDAS and MERRA and SPEI models

It can be noticed that shape of the three-time series curves is similar, but with some slight differences in the values. The best correlation can be observed between models is when analyzing MERRA-2 (the orange line in the figure) and GLDAS (the blue line in the figure) time series. Almost in the whole tested period month amplitudes of the atmospheric energy budget time series are noticed, instead of the period 04.2007-01.2009. But, it need to be said that all three models have a good convergence. For the purpose of the research it was decided to average all three tested models. The result is presented in the figure 2. For the model a trend was computed (dotted line in the figure 2). The trend is almost stable.

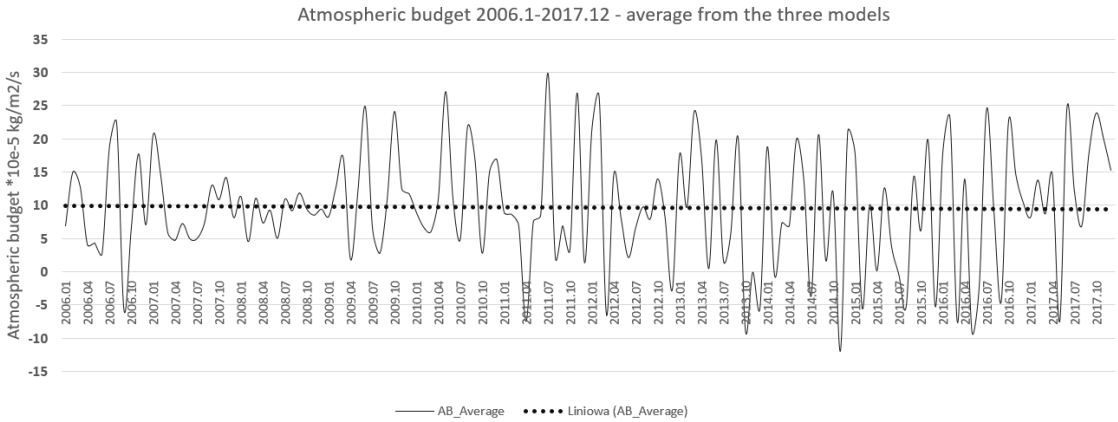


Fig. 2. Averaged time series of the atmospheric energy budget computed on a basis of GLDAS, MERRA-2 and SPEI models with a trend line (dotted)

4.2. Time series forecasting

For the purpose of the time series forecasting a seasonal decomposition was introduced, time series limited to a period 01.2006-12.2015. The time series were limited (24 months is cut), this help me to control the correctness of the thesis of using the ARIMA model. At first, an additive model was introduced, a special case of general linear models. Here, using a linear least-squares fit is calculated for a set of X variables, this let us calculate a dependent Y variable. Based on the additive model, a moving average, and then, corrected time series was estimated. Calculated standard error based on the corrected time series was about 0.08-0.09. The additive model was computed with 15-month delay, the standard deviation of the model was 8.33. The effect of the seasonal decomposition is presented in a figure 3 (seasonality component and corrected time series). Normal probability let us evaluate if the time series can be distributed normally. Estimated values are ranged and presented with comparison to the expected normal value. The result is presented in the figure 4. It can be stated that values are very close to the line of the expected value. I can conclude that the additive model calculated un the research can be used for the prognosis.

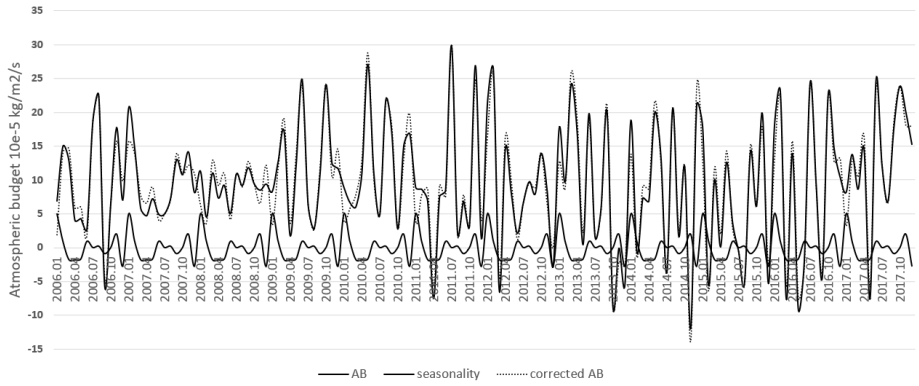


Fig. 3. The atmospheric energy budget time series with added seasonality chart and corrected value oh AB time series

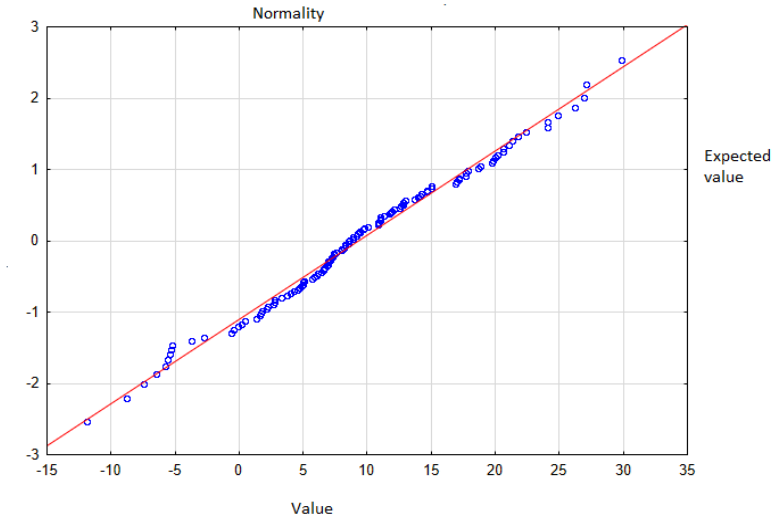


Fig. 4. Chart of the normality (in blue) and expected value (in red) prepared after seasonal decomposition of the atmospheric energy budget time series

For the new time series, a prognosis was prepared with the following parameter: p=2, d=0, q=2 were taken, as gave the best statistics. It means best fitting ARMA (2,0,2) model to the data.

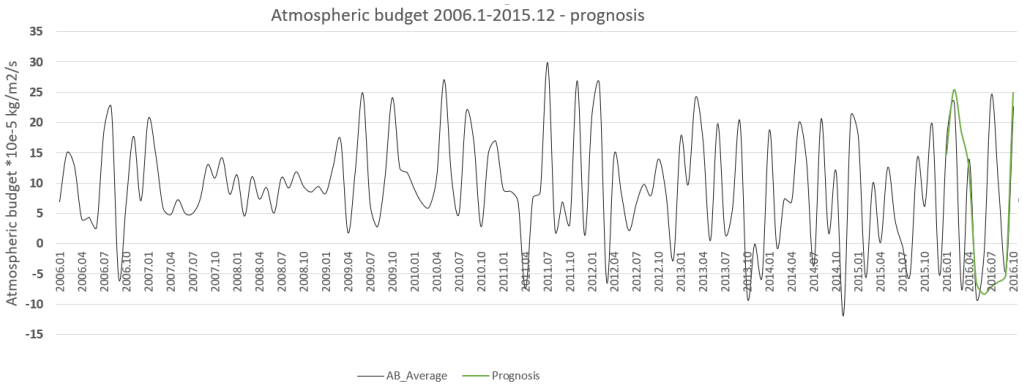


Fig. 5. The atmospheric energy budget time series (in black) with a 10-month prognosis (01.2016-10.2016) prepared with ARIMA model – an attempt; prognosis computed with the time series without seasonality, presented in a figure with seasonality

As a comparison between the real atmospheric energy budget and the prediction for 10 months proved the assumption that the prediction ARIMA can give correct results (values well correlated: 0.74), a prognosis for a period 01.2018-10.2018 was prepared, presented in the figure 6.

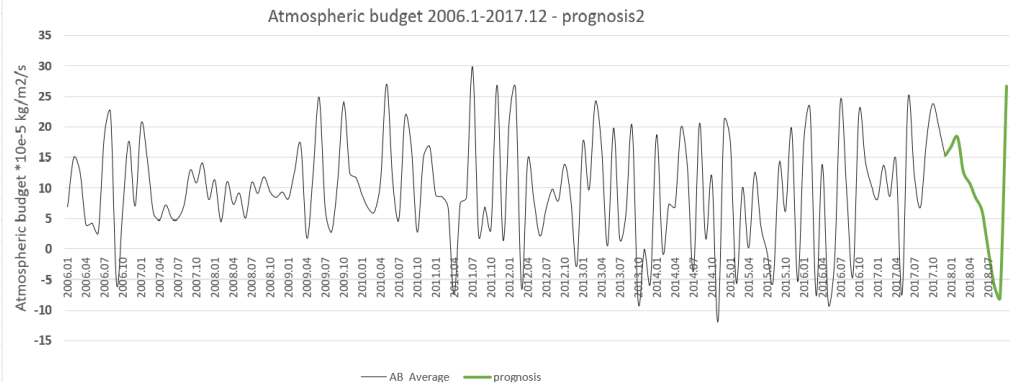


Fig. 6. The atmospheric energy budget time series (in black) with a 10-month prognosis (01.2018-10.2018) prepared with ARIMA model (in green)

5 Conclusions

The aim of the paper was determining the atmospheric energy budget in to the area the Poland. Three models were used: the Global Land Assimilation System (GLDAS), the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2), Standardized Precipitation Evapotranspiration Index (SPEI).

On the basis of the research it was able to notice that the time-series comparisons show strong and positive dependence between the GLDAS and MERRA-2 models, a bit weaker between SPEI and GLDAS, and between SPEI and MERRA-2 models.

Analyzing atmospheric energy budget, it can be said that in Poland summer, winter and spring achieve values above normal (values over zero). During autumn, the atmospheric energy budget is normal (values about zero). Considering energy fluxes, I can say that in Poland usually precipitation values are greater than evapotranspiration.

A comparison between the real atmospheric energy budget and the prediction for 10 months proved the assumption that the prediction ARIMA can give correct results.

6 Acknowledgments

The paper was supported by the National Science Center, UMO-2015/17/ST10/03927 dated from 16.03.2016

7 Bibliography

1. R. Ganguly, D. Kumar, P. Ganguli, G. Short, J. Klausner, "Climate Adaptation Informatics: Water Stress on Power Production",
2. Comput.Sci. Eng. 17, 53–60, DOI: 10.1109/MCSE.2015.106, 2016
3. A. M. Da Silva, C. C. Young, S. Levitus, "Atlas of Surface Marine Data" ,vol.1, Algorithms and Procedures, NOAA Atlas NESDIS 6, U.S. Dep. of Commer., Natl. Oceanic and Atmos. Admin./Natl. Environ. Satellite Data Inf. Serv., Silver Spring, Md, 1994
4. M. Collins, R. Knutti, J. Arblaster, J. L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W. J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A. J. Weaver, M. Wehner, "Long-term climate change: Projections, commitments and irreversibility. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Doschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley", Eds. Cambridge University Press, 1029-1136, DOI: 10.1017/CBO9781107415324.024, 2013
5. M. Byrne, P. O'Gorman, "The Response of Precipitation Minus Evapotranspiration to Climate Warming: Why the "Wet-Get-Wetter, Dry-Get-Drier" Scaling Does Not Hold over Land, Ametsoc, on line, DOI: http://dx.doi.org/10.1175/JCLI-D-15-0369.1, 2015
6. Dingman S.L., 2002, Physical hydrology. Waveland Press, Long Grove, p 646
7. J. Raods, S. Chen, GDAS's GCIP Energy Budgets, 1997, Journal of the Atmospheric Science, DOI: 10.1175/1520-0469
8. S. Swenson, J. Wahr, Estimating Large-Scale Precipitation Minus Evapotranspiration from GRACE Satellite Gravity Measurements, Journal of Hydrometeorology 7(2):252-269, 2006, DOI: 10.1175/JHM478.1
9. Z. Rzepecka, J. Kalita, K. Stepniak, P. Wielgosz, „Time series analysis of radio signals wet tropospheric delays for short-term forecast", Acta Geodynamica et Geomaterialia, 12, 4(180), 2015
10. M. Birylo, K. Pajak, "Statistical approach to the computation of an influence of the Yangtze Dam on gravity fluctuation", Environmental Engineering, eISSN 2029-7092/eISBN 978-609-476-044-0, 2017, DOI: 10.3846/enviro.2017.165

11. M. Birylo, Z. Rzepecka, J. Kuczynska-Siehien, J. Nastula, „Analysis of water budget prediction accuracy using ARIMA models", Water Science and Technology, 2017
12. Narasimhan T.N., 2009, in Encyclopedia of Inland Waters
13. Nau R., "Statistical forecasting: notes on regression and time series analysis", Fuqua School of Bussiness, Duke University, 2017
14. L. Labedzki, B. Bak, E. Kaneck-Geszke, „Wielkość i zmienność ewpotranspracji wskaźnikowej według Penmana-Monteithaw okresie wegetacyjnym w latach 1970-2004 w wybranyc rejonach Polski", Środowisko-obszary wiejskie (IV–VI): t. 12 z. 2 (38) WATER-ENVIRONMENT-RURAL AREAS ISSN 1642-8145s. 159–170, 2012
15. Thornthwaite C. W., 1948, An approach toward a rational classification of climate, Geographical Review. 38 (1): 55–94. doi:10.2307/210739
16. M. Rodell, P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C. J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, D. Toll D, "The Global Land Data Assimilation System", Bull. Amer., 2004
17. B. F. Zaitchik, M. Rodell, F. Olivera, "Evaluation of the Global Land Data Assimilation System using global river discharge data and a source to sink routing scheme", Water Resour. Res., 46, W06507, DOI: 10.1029/2009WR007811, 2010
18. M. Rodell, H. K. Beaudoing, "GLDAS Noah Land Surface Model L4 monthly 1.0 x 1.0 degree V2.0", Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), DOI: 10.5067/QN80TO7ZHFJZ, 2015
19. M. G. Bosilovich, S. Akella, L. Coy, R. Cullather, C. Draper, R. Gelaro, R. Kovach, Q. Liu, A. Molod, P. Norris, K. Wargan, W. Chao, R. Reichle, L. Takacs, Y. Vikhliaev, S. Bloom, A. Collow, S. Firth, G. Labow, G. Partyka, S. Pawson, O. Reale, S. D. Schubert, M. Suarez, "MERRA-2: Initial Evaluation of the Climate, Technical Report Series on Global Modeling and Data Assimilation", 43, DOI: NASA/TM–2015-104606/Vol. 43, 2015
20. M. M. Rienecker, M. J. Suarez, R. Todling, J. Bacmeister, L. Takacs, H. C. Liu, W. Gu, M. Sienkiewicz, D. Koster, R. Gelaro, I. Stajner, J. E. Nielsen, "The GEOS-5 Data Assimilation System—Documentation of versions 5.0.1 and 5.1.0, and 5.2.0.", Tech. Rep. NASA/TM-2008-104606, 101 pp. [Available online at http://gmao.gsfc.nasa.gov/pubs/docs/Rienecker369.pdf.], 2008
21. M. M. Rienecker, M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D. Schubert, L. Takacs, G. K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R. D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C. P. Redder, R. Reichle, F. R. Robertson, A. G. Ruddick, M. Sienkiewicz, J. Woollen, "MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications". J. Climate, 24, 3624–3648, DOI: https://doi.org/10.1175/JCLI-D-11-00015.1, 2011
22. R. G. Allen, L. R. Pereira, D. Raes, M. Smith, "Crop evapotranspiration - Guidelines for computing crop water requirements – FAO", Irrigation and drainage paper 56, 1998
23. M. Birylo, Evaluation of the water cycle determined with atmospheric energy balance for the purpose of surface fluxes monitoring Splitech conference proceedings, 2018

The role of oil prices on the Russian business cycle

Harri Pönkä* and Yi Zheng**

*Ministry of Finance, Finland
**Department of Political and Economic Studies, University of Helsinki

Abstract. We study the role of oil prices in forecasting Russian recession periods with probit models. Our preliminary findings suggest that changes in nominal oil prices are useful predictors of Russian recessions, even when controlling for a number of classic recession predictors. The best in-sample fit is found using a model combining the term spread and the oil price variable, and the pseudo out-of-sample forecasts confirm these findings.

Keywords: Business cycle, Oil prices, Probit model, Recession
JEL classification: C22, G12, G17

1 Introduction

Russia is the second largest producer of natural gas and the third largest producer of oil in the world, with over 109 billion barrels of oil reserves at the end of 2016. Furthermore, it is the largest exporter of oil in the world.[1] The exports of mineral products (consisting mainly of oil and natural gas) accounted for 59.2% of total Russian exports in 2016[2]. Given these figures, it is undeniable that changes in oil and gas prices have a large impact on the economic fluctuations in Russia. In this article, we will analyze the impact of oil price changes on Russian business cycle fluctuations by means of dynamic probit models.

The role of oil prices as a source of business cycle fluctuations has been a topic of wide interest, and was sparked by the two oil crises in the 1970's. Early contributions in the literature include Hamilton (1983), who found statistical evidence of increases in oil prices leading recessions in the US, and since then, the topic has received wide attention (see, e.g., Serletis and Elder (2011) and references therein). Extensions to the literature have suggested that the relationship may be asymmetric (Mork (1989), Hamilton (2011)), as well as dependent

* Disclaimer: The opinions expressed in this article are the author's own and do not reflect the views of the Finnish Ministry of Finance.
** Corresponding author.
[1] Source: BP Statistical Review of World Energy 2017. These figures vary between sources.
[2] Source: Rosstat `http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/figures/activities/`

on whether the shock in oil price is demand or supply driven (Brown and Yucel (2002), Kilian (2009)). Furthermore, the effects of oil price shocks vary between oil producing and importing countries (see, e.g. Mork et al. (1994)), where increases of oil prices have found to have significant positive effects on output in oil exporting countries (see, e.g. Berument et al. (2010)).

In the literature on business cycle fluctuations, binary dependent variable models, such as probit and logit models, have been a standard tool in modelling the probability of recessions since the seminal paper of Estrella and Hardouvelis (1991). The findings based on these models have identified the term spread and stock market returns as useful predictors of recessions (see, e.g., Estrella and Mishkin (1998), Chauvet and Potter (2005), Nyberg (2010), and Ng (2012)) in the US. Further research has suggested that also sentiment (Christiansen et al. (2014)) and credit (Pönkä (2017)) variables have predictive ability for future recession periods.

We contribute to the literature by studying oil price – business cycle relationship in Russia. Although the Russian economy is in many ways dependent on oil production, making it an ideal candidate for research on the topic, the relationship has between oil prices fluctuations and recession periods in Russia not been studied widely in a formal econometric setting. Further motivation for analysing the relationship is given in Figure 1, indicating that three latest recession periods in Russia (as defined in Section 3.1) have coincided with decreases in oil prices. Obviously, there are also other contributing factors to these recessions, discussed e.g. in Smirnov et al. (2017), but the relationship implied by the figure calls for a formal investigation.

The findings suggest that changes in oil prices have predictive ability on future recession periods. Furthermore, models combining the oil price variable with classic recession predictors improve the in-sample performance, as measured with the area under the receiver operating characteristic curve (AUC). The best in-sample fit is found using a model combining the term spread and the oil price variable. The out-of-sample findings are in line with the in-sample results, as models including the term spread and change in oil prices yield the highest AUC:s.

The rest of this paper is organised in the following way. In Section 2, we describe the model and goodness-of-fit measures employed in this paper. In Section 3, we discuss the data, including the business cycle chronology and the explanatory variables. In Section 4, we present the empirical findings of the study. Finally, Section 5 concludes.

2 Empirical Approach

In this section we briefly present the econometric framework and discuss goodness-of-fit measures related to the binary response models.

Fig. 1. Russian recession periods and the Brent oil price 1997Q1–2017Q4

2.1 The probit model

We are interested in understanding the drivers of business cycle fluctuations in Russia, and especially on the role of oil price changes as an explanatory variable. Therefore, throughout the analysis, the dependent variable is the status of the Russian business cycle. In practice, this variable is a binary indicator given by:

$$y_t = \begin{cases} 1, \text{if the economy is in a recession,} \\ 0, \text{if the economy is in an expansion.} \end{cases} \tag{1}$$

As the methodology, we employ probit models using lagged potential predictors, such as changes in oil prices, as explanatory variables. To determine the conditional probability of a recession (p_t), a univariate probit model is specified as

$$p_t = P_{t-1}(y_t = 1) = \Phi(\pi_t), \tag{2}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and π_t is a linear function of the variables in the information set Ω_{t-1}. In the most commonly used model, the so-called static probit model, π_t is specified as

$$\pi_t = \omega + \boldsymbol{x}'_{t-k}\boldsymbol{\beta}, \tag{3}$$

where ω is a constant term and $\boldsymbol{x}_{t-k}$ includes the k:th lagged values of the explanatory variables. We estimate the parameters of the model using maximum likelihood and compute robust standard errors, similarly to Kauppi and Saikkonen (2008).

We also consider an extension to the conventional static probit model. More explicitly, we employ the first-order autoregressive probit model of Kauppi and Saikkonen (2008)

$$\pi_t = \omega + \alpha_1 \pi_{t-1} + \boldsymbol{x}'_{t-k}.\boldsymbol{\beta} \tag{4}$$

An autoregressive structure is introduced into the model by including the lagged value of the linear function π_t. The autoregressive specification of the probit model has been found by e.g. Nyberg (2010, 2014) to outperform static models in predicting US and German recessions.

2.2 Goodness-of-fit Measures

There are several measures to evaluate the goodness-of-fit of binary dependent variable models. The most obvious one is simply the percentage of correct predictions, typically referred to as the success ratio (SR). Formally, a signal forecast for the state of the economy y_t may be defined as

$$\hat{y}_t = \mathbf{1}(p_t > \xi), \tag{5}$$

where the conditional probability of recession p_t is obtained from a probit model, as defined in equation (2). If p_t is larger than a prespecified threshold ξ, we get a signal forecast $\hat{y}_t = 1$ (i.e. recession), and vice versa $\hat{y}_t = 0$ if $p_t \leq \xi$.

In this paper, we employ the threshold $\xi = 0.5$ for SR, which can be seen as natural threshold in (5). However, this is not a fully objective selection, and in some previous studies lower values for ξ have also been used (see,e.g. Nyberg (2010)). The success ratio is certainly important from the practical forecasters point of view, if a sing. However, as recession periods are uncommon compared to expansion periods, the success ratios of relatively uninformative models might turn out high. To test whether the value of the success ratio is higher than that obtained when the realized values y_t and the forecasts $\hat{y}_t$ are independent, we employ the predictability test (PT) of Pesaran and Timmermann (2009).

The Receiver Operating Characteristic (ROC) curve is an alternative method to assess the goodness-of-fit of binary dependent variable models, and it has recently gained popularity in economic applications (see, e.g., Berge and Jorda (2011); Schularick and Taylor (2012); Christiansen et al. (2014)). The ROC curve is a mapping of the true positive rate

$$TP(\xi) = P_{t-1}(p_t > \xi | y_t = 1) \tag{6}$$

and the false positive rate

$$FP(\xi) = P_{t-1}(p_t > \xi | y_t = 0), \tag{7}$$

for all possible thresholds $0 \leq \xi \leq 1$, described as an increasing function in $[0,1] \times [0,1]$ space, with $TP(\xi)$ plotted on the Y-axis and $FP(\xi)$ on the X-axis. A ROC curve above the 45-degree line indicates forecast accuracy superior to a coin toss. Given that it takes into account all possible thresholds ξ, the ROC curve is a more robust method to evaluate the goodness-of-fit of a model than the success ratio.

The information in the ROC curve is typically summarized by the area under the ROC curve (AUC), which is simply the integral of the ROC curve between zero and one. Therefore, the AUC also gets values between 0 and 1, with the value of 0.5 corresponding a coin toss and the value 1 to a perfect forecast. Any improvement over the AUC=0.5 indicates statistical predictability. We test the null hypothesis of AUC= 0.5 implying no predictability using standard techniques (see Hanley and McNeil, 1982).

Another commonly used measure is the pseudo-R^2 of Estrella (1998), which is a counterpart of the coefficient of determination (R^2) designed for binary response models. The measure is given by

$$psR^2 = 1 - \Big(\frac{logL_u}{logL_c}\Big)^{-(2/T)logL_c}, \tag{8}$$

where $logL_u$ and $logL_c$ are the maximum values of the constrained and unconstrained log-likelihood functions respectively, and T is the sample size. This measure takes on values between 0 and 1, and can be interpreted in the same way as the coefficient of determination in the usual linear predictive regression model. In Section 4, we also report the adjusted form of (8) (see Estrella (1998)) that takes into account the trade-off between improvement in model fit and the use of additional estimated parameters.

3 Data

In this section, we discuss the data employed in this study. The sample used in the study is 1997–2017 and the data is quarterly.

3.1 The Russian Business Cycle

One of the key issues in terms of data is the selection of the business cycle chronology, as defined in equation (1). Unlike in the U.S., where the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER)[3] determines the official turning points, in Russia there is no such official chronology of recessions and expansions. However, there are a number of ways to determine the turning points based on data. Smirnov et al. (2017) recently established a monthly reference chronology for the Russian economic cycle from the early 1980s to mid-2015, using various seasonal adjustment methods and dating methods. In this paper, we define the turning points for business cycles using the

[3] http://www.nber.org/cycles/recessions.html

Bry-Boschan (BB) algorithm (Bry and Boschan (1971)), which is a commonly used method in the literature. The dating is based on the algorithm used for seasonally adjusted real GDP data for the period 1997Q1-2017Q4. The resulting chronology is presented in Table 1 and was plotted with the Brent oil price in Figure 1.

Table 1. Turning points for the Russian Business Cycle

Peaks	Troughs
1997Q4	1998Q3
2008Q2	2009Q2
2014Q3	2016Q3

The BB algorithm finds three recession periods in the period 1997Q1-2017Q4. These findings are in line with those of Smirnov et al. (2017), who use monthly data in their reference chronology.

3.2 Predictive Variables

The oil price variable selected for the study is the Brent oil price in US dollars. The main specification used is the quarterly change in prices (DOIL).[4] As we are interested in studying the predictive ability of oil prices over and above other predictors, we employ a number of commonly used predictors of recessions as control variables. Several studies on other countries have suggested that financial variables are useful predictors of real activity and recessions (see, e.g., Stock and Watson (2003)). Among the most useful financial leading indicators are the term spread (TS) and stock returns (RET) (see, e.g., Estrella and Mishkin (1998) and Nyberg (2010)). Therefore, these predictors are obvious choices as additional predictors. Along with these variables, also the short term interest rate has been employed as a predictor in a number of studies (see, e.g. Wright (2006) and Pönkä (2017)).

Sentiment variables, such as consumer confidence indices, are a particularly interesting group of variables, due to their forward-looking nature. Christiansen et al. (2014) found that the consumer confidence and purchasing managers' indices are useful predictors of US recession periods, even when combined with classic recession predictors and common factors based on a large panel of economic and financial variables. Based on these findings, we include the consumer confidence index (CCI) in our set of potential predictors.

In Table 2, the predictive variables have been listed, along with the abbreviations and the starting points of the sample for each variable. Altogether, some of the data ar already available from the beginning of 1997, so we are able to

[4] In later versions of the paper, we also plan to use four-quarter changes of oil prices. As a robustness check, we also plan to test the findings using the Urals oil prices.

Table 2. Sample starting point for the leading indicators

Variable	Abbreviation	Starting point
First difference of the Brent oil price	$DOIL_t$	1997Q1
First difference of the three-month interest rate	DTM_t	1997Q2
Logarthmic return of the stock index	RET_t	1998Q1
First difference of the consumer confidence index	$DCCI_t$	1999Q1
Term spread (10y bond yield minus the three-month interest rate)	TS_t	2001Q3

Notes: This table presents the predictive variables and their starting points.

include the first recession period (1997Q4-1998Q3) in the sample. On the other hand, the term spread could only used starting from 2001Q3 onwards.[5]

Furthermore, we present the correlations between the predictive variables in Table 3. The highest correlations are found for the change in consumer confidence. It is positively correlated with the term spread and the change in oil prices and negatively correlated with the change in the short term interest rate.

Table 3. Correlations between employed predictive variables

	$DOIL_t$	DTM_t	RET_t	$DCCI_t$	TS_t
$DOIL_t$	1	-0.449	0.441	0.577	0.482
DTM_t		1	-0.396	-0.519	-0.401
RET_t			1	0.362	0.301
$DCCI_t$				1	0.541
TS_t					1

Notes: This table presents the correlation coefficients between the employed predictive variables.

4 Results

In this Section, we present the main findings of our research. We first study the performance of the individual explanatory variables as predictors of the Russian business cycle. We allow each predictor to have a lag length beween one to four quarters, as findings from previous literature has suggested that different variables have predictive ability at different lag lengths.

The findings in Table 4 illustrate that changes in oil prices do have predictive ability on future recession periods (using first and second lags). However, it is the term spread that performs clearly the best as a predictor. The model including the first lag of the term spread yields an AUC of 0.969, whereas the one with the change in oil price yields an AUC of 0.743, which is also relatively high for a single predictor. The difference in fit is only partly explained by the shorter

[5] It is available starting from 1999Q1, but at that point the ten-year government bond yields were so high in the first two years of the sample that using the data from that point onwards caused problems the estimation of the model.

Table 4. In-sample results for single-predictor probit models

	Variable	Coeff.	adj.psR^2	BIC	SR	AUC
First lags						
1	$DOIL_{t-1}$	-0.052**	0.101	39.845	0.841	0.743***
2	DTM_{t-1}	0.169***	0.092	40.185	0.841	0.703***
3	RET_{t-1}	-0.022***	0.051	39.812	0.810*	0.683**
4	$DCCI_{t-1}$	-0.101***	0.107	34.022	0.880	0.689**
5	TS_{t-1}	-0.806***	0.557	18.156	0.908	0.969***
Second lags						
6	$DOIL_{t-2}$	-0.048*	0.087	40.193	0.840	0.720***
7	DTM_{t-2}	0.091*	0.015	43.033	0.815	0.613*
8	RET_{t-2}	-0.023**	0.055	37.966	0.821	0.728***
9	$DCCI_{t-2}$	-0.069***	0.041	36.211	0.838	0.649**
10	TS_{t-2}	-0.240**	0.204	29.108	0.828	0.903***
Third lags						
11	$DOIL_{t-3}$	-0.012	Neg.	43.980	0.800	0.628*
12	DTM_{t-3}	-0.012	Neg.	44.385	0.800	0.473
13	RET_{t-3}	-0.008	Neg.	38.896	0.831	0.623*
14	$DCCI_{t-3}$	-0.043	Neg.	37.514	0.822	0.642*
15	TS_{t-3}	-0.154	0.085	32.602	0.778**	0.858***
Fourth lags						
16	$DOIL_{t-4}$	-0.003	Neg.	42.746	0.810	0.517
17	DTM_{t-4}	0.010	Neg.	42.747	0.810	0.570
18	RET_{t-4}	-0.005	Neg.	38.990	0.829	0.600
19	$DCCI_{t-4}$	-0.047	0.004	37.121	0.806	0.673**
20	TS_{t-4}	-0.143	0.072	32.791	0.774	0.829***

Notes: This table presents the findings from single-predictor probit models for Russian recessions. The table includes findings for the oil price and control variables. Robust standard errors are given in brackets (see Kauppi and Saikkonen (2008)). The goodness-of-fit measures are described in detail in Section 2. In the table, *, **, and *** denote the statistical significance of the estimated coefficients and the AUC at 10%, 5% and 1% significance levels, respectively.

sample used for the term spread. In line with previous findings in the literature, the term spread has predictive ability even using longer lag lengths.

The findings from the single-predictor models were rather promising. Following the typical convention, we proceed by estimating multiple predictor models. Moreover, we estimate models combining the oil price variable with each of the other predictors. We allow each variable to have a lag between one and four, and report the best performing models in Table 5.

Model 21 uses the combination of oil prices and the short term interest rate. The AUC is 0.797, which is higher than for the individual predictors in Table 4, but lower than for Models 22–24. In Models 21–23, the coefficients of the oil price variable are statistically significant. Model 24, including the term spread, yields the highest AUC among the two-predictor models (0.974). This is even higher than for the single-predictor model (Model 5). However, the coefficient for the oil price variable is no longer statistically significant. Overall, we find that models combining the oil price variable with classic recession predictors yield stonger results than single-predictor models.

Table 5. Estimation results for in-sample predictive models

Variable	21	22	23	24
$DOIL_{t-2}$		-0.067**	-0.046*	
		(0.031)	(0.027)	
$DOIL_{t-3}$	-0.054***			0.024
	(0.021)			(0.016)
DTM_{t-1}	0.307***			
	(0.064)			
RET_{t-1}		-0.051***		
		(0.016)		
$DCCI_{t-3}$			-0.041	
			(0.033)	
TS_{t-1}				-0.723***
				(0.168)
CONST	-1.035***	-0.871***	-0.948***	-0.227
	(0.276)	(0.293)	(0.279)	(0.259)
adj.psR^2	0.214	0.213	0.106	0.327
BIC	35.345	32.107	35.275	26.909
SR	0.835	0.882**	0.847	0.857***
AUC	0.797***	0.812***	0.812***	0.974***

Notes: This table presents the findings from probit models for Russian recessions. In the table, *, **, and *** denote the statistical significance of the estimated coefficients, the Pesaran and Timmermann (2009) (PT) predictability test for the success ratio, and the AUC at 10%, 5% and 1% significance levels, respectively. See also notes to Table 4.

4.1 Out-of-sample Findings

As previous forecasting literature has shown, good in-sample fit does not necessarily imply good out-of-sample performance. Therefore, in this section, we will examine the pseudo out-of-sample forecasting performance of our models. We use an expansive window forecasting approach with estimation samples ranging from 2001Q3–2009Q4 to 2001Q3–2017Q3 and report the results of one- and two-quarter-ahead forecasting horizons.[6] Therefore, in our forecasting sample (2010Q1–2017Q4), there is only one recession. This limitation is due to the small number of recessions in the full sample, as for each estimation sample we need at least one recession period.

Table 6. Out-of-sample results for models including credit variables and classic predictors

Forecast horizon: 1 quarter									
Model	DOIL	DTM	RET	DCCI	TS	21	22	23	24
SR	0.813	0.781	0.438	0.781	0.875	0.750	0.563	0.719	0.813
AUC	0.787***	0.792***	0.237	0.783***	0.932***	0.778***	0.662*	0.918***	0.932***
Forecast horizon: 2 quarters									
SR	0.719	0.781	0.563	0.719	0.719	0.688	0.469	0.719	0.688
AUC	0.894***	0.758***	0.522	0.783***	0.831***	0.676**	0.261	0.836***	0.609

Notes: This table presents the one-to-four-quarter-ahead forecasting results from static probit models for Russian recession periods. See also the notes to Table 4.

[6] This selection is based on the sample length available for the term spread.

The findings indicate that the models including the term spread (TS and Model 24) perform best among the models in one-quarter-ahead forecasts, with out-of-sample AUC:s of 0.932 for both models. Model 23, including the oil price and consumer confidence variables, also perform relatively well (AUC=0.918). In the case of two-quarter-ahead forecasts, the single predictor model including the oil price variable performs the best, yielding an AUC of 0.894, whereas the model including the term spread performs the second best (AUC=0.831).

5 Conclusion

In this paper, we have studied the role of oil prices on Russian business cycle fluctuations. The findings indicate that changes in oil prices do have predictive ability on future recession periods. However, the term spread, defined as the difference between the ten-year government bond and the three-month interest rate, yields stronger results based on the area under the ROC curve (AUC), which which has been the main measure of interest in this paper. Models combining the changes in oil prices with classic recession predictors improve the in-sample performance of the models. In terms of pseudo out-of-sample results, models including the term spread perform the best in one-quarter-ahead forecasts, whereas a model including the change in oil prices, yields the highest AUC in two-quarter-ahead forecasts.

The findings of this paper could be extended in a number of ways. In further versions, we will include findings based on autoregressive probit models. A further extension would be to use of a larger set of variables. In this paper, we have considered the predictive ability of oil prices and classic recession predictors, but the use of common factors based on a large panel of financial and macroeconomic variables could also be considered, in the lines of Christiansen et al. (2014) and Pönkä (2017).

Bibliography

T.J. Berge and O. Jorda. Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics*, 3: 246–277, 2011.

H. Berument, N.B. Ceylan, and N. Dogan. The impact of oil price shocks on the economic growth of selected MENA countries. *Energy Journal*, 31:149–176, 2010.

S.P.A. Brown and M.K. Yucel. Energy prices and aggregate economic activity: An interpretative study. *Quarterly Review of Economics and Finance*, 42: 193–208, 2002.

G. Bry and C. Boschan. *Evaluating the Classification of Economic Activity into Recessions and Expansions*. NBER Books. National Bureau of Economic Research, Inc, 1971.

M. Chauvet and S. Potter. Forecasting recessions using the yield curve. *Journal of Forecasting*, 24:77–103, 2005.

C. Christiansen, J.N. Eriksen, and S.T. Møller. Forecasting US recessions: The role of sentiment. *Journal of Banking and Finance*, 49:459–468, 2014.

A. Estrella. A new measure of fit for equations with dichotomous dependent variables. *Journal of Business and Economic Statistics*, 16:198–205, 1998.

A. Estrella and G.A. Hardouvelis. The term structure as a predictor of real economic activity. *Journal of Finance*, 46:555–576, 1991.

A. Estrella and F.S. Mishkin. Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80:45–61, 1998.

J.D. Hamilton. Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91:228–248, 1983.

J.D. Hamilton. Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic Dynamics*, 15:364–378, 2011.

J.A. Hanley and B.J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143:29–36, 1982.

H. Kauppi and P. Saikkonen. Predicting U.S. recessions with dynamic binary response models. *Review of Economics and Statistics*, 90:777–791, 2008.

L. Kilian. Not all oil shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99:1053–1069, 2009.

K.A. Mork. Oil and the macroeconomy when prices go up and down: An extension of Hamilton's results. *Journal of Political Economy*, 97:740–744, 1989.

K.A. Mork, O. Olsen, and H.T. Mysen. Macroeconomic response to oil price increases and decreases in seven OECD countries. *Energy Journal*, 15:19–35, 1994.

E.C.Y. Ng. Forecasting US recessions with various risk factors and dynamic probit models. *Journal of Macroeconomics*, 34:112–125, 2012.

H. Nyberg. Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting*, 29:215–230, 2010.

H. Nyberg. A bivariate autoregressive probit model: Business cycle linkages and transmission of recession probabilities. *Macroeconomic Dynamics*, 18:838–862, 2014.

M.H. Pesaran and A. Timmermann. Testing dependence among serially correlated multi-category variables. *Journal of the American Statistical Association*, 485:325–337, 2009.

H. Pönkä. The role of credit in predicting U.S. recessions. *Journal of Forecasting*, 36:221–247, 2017.

M. Schularick and A.M. Taylor. Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review*, 102:1029–1061, 2012.

A. Serletis and J. Elder. Introduction to oil price shocks. *Macroeconomic Dynamics*, 15:327–336, 2011.

S.V. Smirnov, N.V. Kondrashov, and A. V. Petronevich. Dating cyclical turning points for russia: Formal methods and informal choices. *Journal of Business Cycle Research*, 13:53–73, 2017.

J.H. Stock and M.W. Watson. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41:788–829, 2003.

J.H. Wright. The yield curve and predicting recessions. *Finance and Economics Discussion Series*, 7, 2006.

Seasonal Variations of Sea Level in the Polish Coastal Zone from Satellite Altimetry and Tide Gauge Data

Katarzyna Pajak[1)], Monika Birylo[1)], Joanna Kuczynska-Siehien[1), 2)], Kamil Kowalczyk[1)]

[1)] Faculty of Geodesy, Geospatial and Civil Engineering, University of Warmia and Mazury in Olsztyn, Oczapowskiego ~~St.~~ 2, Olsztyn, Poland
[2)] ~~University of Calgary,~~ Department of Geomatics Engineering, University of Calgary, 2500 University Dr. NW, Calgary, Canada ~~, University of Warmia and Mazury in Olsztyn, Faculty of Geodesy, Geospatial and Civil Engineering~~

Sformatowano: Kolor czcionki: Czarny

Abstract

The seasonal sea level variations have been determined at five locations in the Polish coastal zone from satellite altimetry and tide gauge water level data for the period 1993-2015. In order to analyze the dynamics of change in sea level, the linear trend was estimated from satellite altimetry and tide gauges time series. The real data time series and the data without seasonal variations were used. The correlation between the time series from different data sets have been determined. The seasonal indicators were used to estimate the seasonal variations of changes of sea level. The absolute indicators showed that, as a consequence of the influence of seasonal factors, the values of the given phenomenon are lower or higher than those estimated with the linear trend function. Baltic Sea level shows a mean annual cycle that usually peaks in the Autumn (September, October, November and the Winter months (December and January) and attains its minimum at the end of Spring (May). Despite the fact that the two considered kinds of measurements differ, since satellite altimetry integrates sea level anomalies and tide gauges observe real time water level, it is obvious that these phenomena are strictly connected with each other. The obtained mean correlation coefficient between data sets from satellite altimetry and tide gauges is 0.92.

Introduction

The seasonality of sea level changes is a variable which occurs in sea level anomalies in relation to seasons, in which various phenomena that have an essential influence on the changes appear. Seasonal changeability of the sea is connected mainly with meteorological factors, water balance, evaporation and changes in sea topography, or changes in water density (Chen et al. 2005). Therefore, accurate estimation of the sea level change is important to evaluate and predict its impact on coastal regions.

Since the early 1990s, sea level is routinely measured with quasi-global coverage and a few days/weeks revisit time (called “orbital cycle") by altimeter satellites. The global sea level change has also been measured from numerous networks of coastal tide gauges around the world since the

18th century (Douglas 2001). Tide gauge data is important for determining global or local sea level rise with respect to a global geocentric reference frame (Deo et al. 2013). Tide gauges measure relative sea level, which is directly affected by corresponding ~~ground~~ land motions (Rocco 2016). In contrast to tide gauges, satellite altimetry measures sea level variations relative to the center of mass of the Earth. Coastal sea level variability is therefore the most challenging to measure and interpret (Passaro et al. 2015). Sea level change analysis are important factors used for coastal area. Moreover, sea level change modeling is used widely to evaluate and study shoreline and climate changes (Kaloop et al. 2016).

In the analyses of seasonal variations changes in sea level one can apply the seasonal indicators method. In the method, the seasonal indicators for particular phases of cycles are determined. In the article the authors made the analysis of seasonal fluctuations in the Baltic Sea level in selected stations distributed in Polish coastal zone of the Baltic Sea.

In this paper we used an additive model (Fase et al., 1973). In the additive model the seasonal, cyclical and random variations are absolute deviations from the trend (they do not depend on the level of the trend). The reasons for the occurrence of seasonal fluctuations are mostly natural. It means that their occurrence is closely connected with the course of the following seasons and also with changes in climate conditions. The choice of a given method for distinguishing seasonal fluctuations, depends first of all on the course of the general linear trend estimation (particularly on the type of function that describes the tendency), and on types of seasonal fluctuations (Australian Bureau of Statistics, 2018).

In the indicators method used in this paper, one can distinguish ~~four~~ two stages of work:

- the computation of the linear trend model,

- the elimination of the linear trend estimation from the time series (in the case of the additive model, it is achieved by estimating the difference between the empirical value of the analyzed variable and the theoretical value gained from the linear trend estimation model; the gained values are independent from trends, but they contain seasonal and random fluctuations) (Ostasiewicz et al. 2008).

In this paper, we estimated seasonal indicators for the investigation of the seasonal variations of the sea level in the Polish coastal zone. For the purpose of seasonal adjustment, the time series ~~was~~ were assumed to be observed monthly and consisted of distinct elements: the trend, the seasonal, cyclical and irregular components. The objective of seasonal adjustment was to identify and estimate the seasonal effects, and to remove them from the time series.

2 Observation data

2.1. Satellite altimetry

The global merged Maps of Sea Level Anomaly grid data from the Copernicus Marine and Environment Monitoring Service (CMEMS) were used. This product is processed by the SL-TAC multimission altimeter data processing system. It processes data from all altimeter missions: Jason-3, Sentinel-3A, HY-2A, Saral/AltiKa, Cryosat-2, Jason-2, Jason-1, T/P, ENVISAT, GFO, ERS1/2. The altimetric measurements have been corrected for atmospheric effects (ionospheric delay and dry/wet tropospheric effects) and geophysical processes (solid, ocean, and pole tides, loading effect of the ocean tides, sea state bias, and the Inverted Barometer response of the ocean). In order to investigate the seasonal variations of the sea level in the Polish coastal zone, the mean monthly satellite altimetry data set is obtained from the daily data at the closest points to the TG location.

2.2. Tide gauge

The tide gauge (TG) at the coast can measure relative sea level variations with respect to the Earth's crust (Woodworth and Player, 2003; Kowalczyk, 2006). In the study, the 5 tide gauges stations data with the time span 23 years from 1993 to 1999 was acquired from the Permanent Service for Mean Sea Level (PSMSL) global network of tide gauges (Fig. 1) (PSMSL, see http://www.psmsl.org and Woodworth and Player 2003). The data for the period from 1999 to 2015 was received from the Institute of Meteorology and Water Management National Research Institute, Poland. The time series of monthly TG data are used to analyze the relative sea level changes (Piecuch, 2016).

3. Results and discussion

The sea level change time series were obtained from satellite altimetry and tide gauge data. Fig. 1 shows the sea level variations at the all analyzed stations.

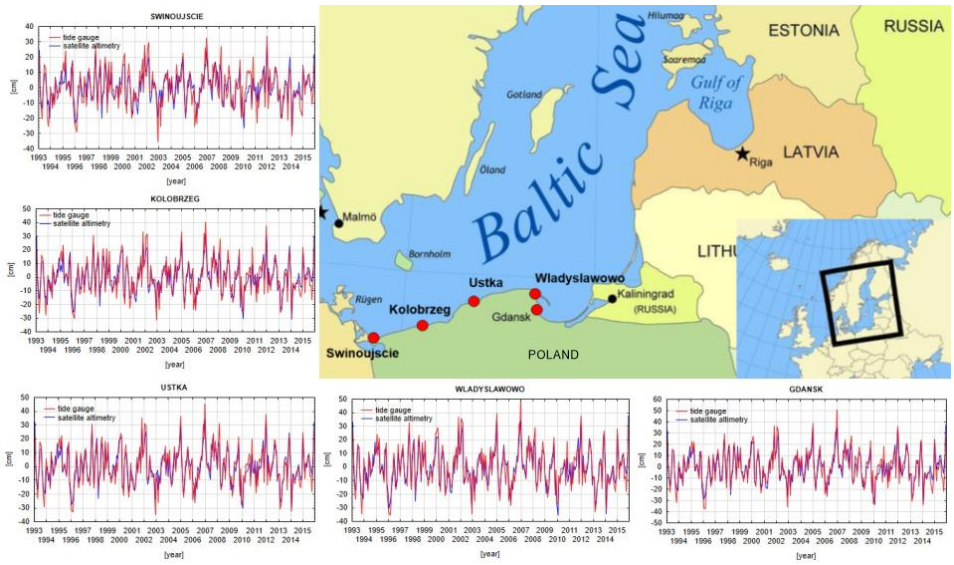


Fig 1. The location of co-located tide gauge stations used in the study with the un-trended sea level change time series ~~(with trend removed)~~ from satellite altimetry- and tide gauge at all analyzed stations

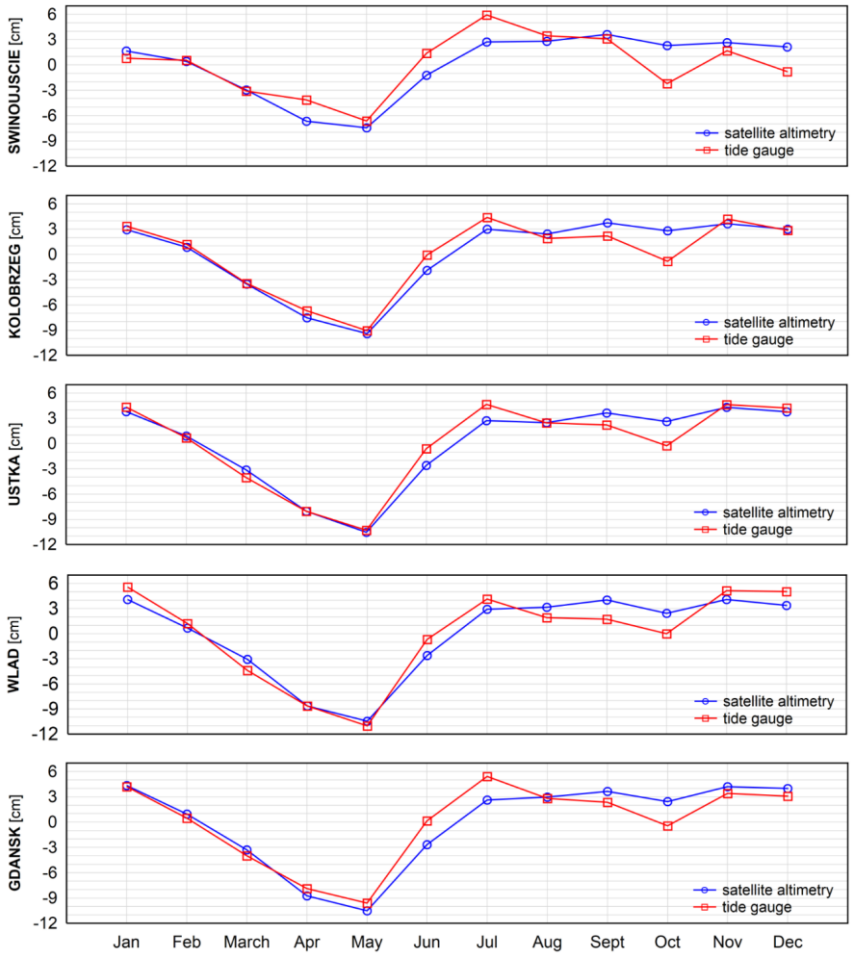
~~It is seen that the sea level change time series have a strong seasonal signal and long-term trend, which agree remarkably well for two independent techniques.~~ The satellite altimetry time series and tide gauge water level data series have been analyzed at the same time period (Table 1). Additionally, at each station, the correlation coefficient and the coefficient of determination between the satellite altimetry and the tide gauge data were computed (Taylor 1982).~~) (Table 1).~~ It is seen (Figure 1) that the sea level change time series have a strong seasonal signal and long-term trend, which agree remarkably well for the two independent techniques.

Table 1. Trends, correlations and coefficients of determination *(R^2)* of sea level change from satellite altimetry and tide gauge data at the same observation period (1993-2015)

Stations	Trend (mm/yr)		Correlation coefficient	R^2
	Satellite altimetry	Tide gauge		
Swinoujscie (53.917ºN, 14.233ºE)	3.1 ± 0.9	3.3 ± 1.1	0.88	0.77
Kolobrzeg (54.183 ºN, 15.550 ºE)	3.6 ± 1.0	3.2 ± 1.2	0.92	0.85
Ustka (54.583 ºN, 16.867 ºE)	3.8 ± 1.1	3.7 ± 1.3	0.94	0.89
Wladyslawowo (54.800 ºN, 18.417 ºE)	3.8 ± 1.2	1.3 ± 1.4	0.93	0.86
Gdansk (54.400 ºN, 18.683 ºE)	3.8 ± 1.2	1.6 ± 1.4	0.93	0.87

Linear trends for satellite altimetry and tide gauge measurements were evaluated over the same period at each station. According to Table 1, the differences of linear trend between tide gauge and altimeter data were considered small. However, at Wladyslawowo and Gdansk stations these differences appear larger, which may be due to the larger movements of the Earth's crust. The mean correlation coefficient between satellite altimetry and tide gauges is 0.92. Consequently, it can be stated that there is a high positive correlation between satellite altimetry and tide gauge.

In the aim of analysis of seasonal fluctuations, we used the seasonal indicators method. We computed the absolute fluctuations, as the fluctuation amplitudes in corresponding phases of a cycle were similar, by estimating the difference between the empirical value of the analyzed variable and the theoretical value obtained from the linear trend estimation model. The obtained values were independent of trends, but they contain seasonal and random fluctuations. The absolute indicators and the sum of all seasonal indicators were estimated for the tide gauge data and sea level anomaly. In the case of the absolute indicators, the sum coincided with the expected value (Ostasiewicz et al., 2008), i.e. it was 0. Figure 2 presents the calculated seasonal indicators in relation to the estimated trend value. The indicators in the additive model show how much the level of the phenomenon is higher or lower in a given month than the estimated trend function.



In September, November and January the sea level changes (Fig. 2) are the highest (they are higher than the estimated trend function) as a result of seasonal indicators activity from the satellite altimetry data. Moreover, in July and January the sea level changes are the highest from the tide gauge data. The sea level changes from both data sets are the lowest in May in almost all studied stations -at the Polish coastal zone-.

The sea level change determined from satellite altimetry and the fitted seasonal variations are shown in Fig. 3, and the results are collected in Table 2. The correlations show successful fitting for both

data sets. The adjustment of the satellite altimetry time series yields the correlation value 0.41 (seasonal curves with absolute indicators). The tide gauge time series and the fitted seasonal curves are shown in Figure 4 and summarized in Table 2. The adjustment of the tide gauge time series yields 0.35 (seasonal curves with absolute indicators).

Table 2. Results of the satellite altimetry and tide gauge time series fitting~~,~~ determined for the period 1993-2015.

Station	Correlation coefficient	
	Satellite altimetry	Tide gauge
Swinoujscie (53.917ºN, 14.233ºE)	0.42	0.33
Kolobrzeg (54.183 ºN, 15.550 ºE)	0.43	0.35
Ustka (54.583 ºN, 16.867 ºE)	0.43	0.38
Wladyslawowo (54.800 ºN, 18.417 ºE)	0.40	0.35
Gdansk (54.400 ºN, 18.683 ºE)	0.42	0.32

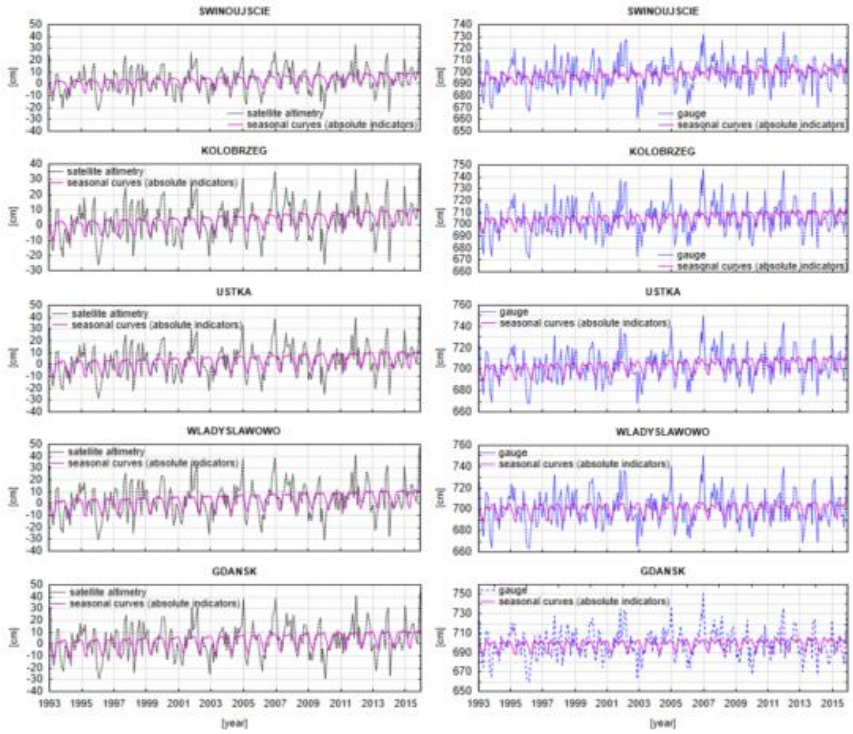


Figure 3. Satellite altimetry [data ~~-~~](#) and tide gauge data~~-~~ [](#) and the best fitting seasonal variation curves

Sformatowano: Do lewej

(pink line) at the all analyzed stations.

Moreover, we calculated the correlation coefficients between two independent time series (seasonal curves). Table 3 shows the correlations between the seasonal curves (absolute indicators) from satellite altimetry and from tide gauges data. The correlation of seasonal variations from two independent observation techniques was 0.93, which is a very high linear correlation between two time series.

Table 3. Correlation coefficients between satellite altimetry and tide gauge; results for seasonal indicators.

Station	Indicators				Correlation coefficient	R^2
	Satellite altimetry		Tide gauge			
	Max	Min	Max	Min		
Swinoujscie (53.917ºN, 14.233ºE)	Sept	May	July	May	0.81	0.66
Kolobrzeg (54.183 ºN, 15.550 ºE)	Sept	May	July	May	0.95	0.90
Ustka (54.583 ºN, 16.867 ºE)	Nov	May	July	May	0.96	0.92
Wladyslawowo (54.800 ºN, 18.417 ºE)	Nov	May	Jan	May	0.96	0.92
Gdansk (54.400 ºN, 18.683 ºE)	Jan	May	July	May	0.95	0.90

We found that there is a high positive dependence between seasonal curves from satellite altimetry and seasonal curves from tide gauge. The mean R^2 coefficient is 0.86.

4. Summary

In our paper we estimated the seasonal indicators to investigate the seasonal variations of the sea level change from satellite altimetry and tide gauge data. We removed the seasonal components from the time series. Many approaches for removing the seasonal variations can be taken. We used the seasonal indicators to estimate the dynamics of changes of sea level.

One type of models, namely the additive model, have been used. In the additive model, the seasonal variation is independent of the absolute level of the time series, but it takes approximately the same magnitude each year.

The observations of the graphs prove that the authors gained the good fitting from the linear trend function and the absolute seasonal indicators.

Based on ~~these~~ the five analyzed~~5~~ stations, the obtained mean annual amplitude in the sea level anomalies time series ~~at five stations~~ is 4.02 cm, in the tide gauge level data the mean annual amplitude is 5.19 cm.

The time series of satellite altimetry and tide gauge data present almost similar behaviour in the sea

level change. The differences of linear trend between tide gauge and altimeter data were considered small. However, at Wladyslawowo and Gdansk stations these differences appear larger, which may be due to the larger movements of the Earth's crust or due to reasons related to human activity.

References

Australian Bureau of Statistics. 2018. http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics. Last accessed 21 May 2018.

Deo M.N, Govind R., El-Mowafy A. 2013. The Stability of Tide Gauges in the South Pacific Determined from Multiepoch Geodetic Levelling, 1992 to 2010. *Marine Geodesy*, Vol. 36, Iss. 3, http://dx.doi.org/10.1080/01490419.2013.786003.

Donner R.V., EhrckeR., Barbosa S.M., Wagner J., Donges J.F., Kurths J. 2012. Spatial patterns of linear and nonparametric long-term trends in Baltic sea- level variability, NonlinearProc. *Geoph*., 19(1), 95–111, http://dx.doi.org/10. 5194/npg-19-95-2012.

Douglas BC. 2001. Sea level change in the era of the recording tide gauge. In: Douglas BC, Kearney MS, Leatherman SP, editors. Sea level rise, history and consequences. *International Geophysics Series*, vol. 75. Academic Press. 2001. p. 37e64.

Fase, M. M. G., Koning, J., Volgenant, A. F. (1973). An Experimental Look at Seasonal Adjustment. De Economist, 121(5), pp. 441–480.

Kaloop Mosbeh R., Rabah M., Elnabwy M. 2016. Sea Level Change Analysis and Models Identyfication Based on Short Tidal Gauge Measurements in Alexandria, Egypt. *Marine Geodesy* Vol. 39, Iss. 1, http://dx.doi.org/10.1080/01490419.2015.1134735.

Kowalczyk, K. 2006. New model of the vertical crustal movements in the area of Poland. *Geodezija ir Kartografija*, *32*(4), 83-87.

Ostasiewicz S., Rusnak Z., Siedlecka U. 2008. Statistics. Elements of Theory and Tasks. *Publisher Economic Academy in Wroclaw.*

Passaro, M., P. Cipollini, and J. Benveniste, 2015: Annual sea level variability of the coastal ocean: The Baltic Sea–North Sea transition zone. J. Geophys. Res. Oceans, **120**, 3061–3078, doi:https://doi.org/10.1002/2014JC010510.

Permanent Service for Mean Sea Level. 2018: Tide gauge data. Permanent Service for Mean Sea Level, accessed 26 May 2018. [Available online at http://www.psmsl.org/data/obtaining/.]

Piecuch Christopher G. 2016. Annual Sea Level Changes on the North American Northeast Coast: Influence of Local Winds and Barotropic Motions. American Meteorological Society. Journals online https://doi.org/10.1175/JCLI-D-16-0048.1.

PSMSL. 2018. Permanent Service for Mean Sea Level: Tide Gauge Data. Available at

http://www.psmsl.org. Last accessed 21 May 2018.

Rocco, Francois Vincent. 2016. Sea level trends in the Mediterranean from tide gauges and satellite altimetry. [Laurea magistrale], Università di Bologna, Corso di Studio in Fisica del sistema terra [LM-DM270].

Stammer D., Cazenave A., Ponte R.M., Tamisiea M.E. 2013. Causes for contemporary regional sea level changes. *Annu Rev Mar Sci,* 5 (2013), pp. 21–46 http://dx.doi.org/10.1146/annurev-marine-121211-172406.

Taylor R. 1990. Interpretation of the Correlation Coefficient: A Basic Review. Journal of Diagnostic Medical Sonography (JDMS). Volume: 6 issue: 1, page(s): 35-39, https://doi.org/10.1177/875647939000600106.

Woodworth, P.L., Player, R. 2003. The permanent service for mean sea level: an update to the 21st century. J. Coastal Res. 19, 287–295.

The Performance of the Wavelet Halt-Winters Hybrid Model in Forecasting the Groundwater Level Time Series (Case Study: Urmeih Coastal Aquifer, Iran)

H.R. Nassery[a], A. Mirarabi[a], M. Nakhaei[b], F. Alijani [a]

[a] Department of mineral geology and hydrogeology, Faculty of Earth science, Shahid Beheshti University, Tehran, Iran
[b] Department of Applied Geology, Faculty of Earth Sciences, Kharazmi University, Tehran, Iran
H-nassery@sbu.ac.ir

Abstract:

For management and planning valuable groundwater resources, it is very important to predict groundwater level and understand correctly the process of aquifer changes. In this paper for the first time, the wavelet Halt-Winters hybrid models (WHW) have been explored and tested for Groundwater forecasting. For this purpose, a 16-year dataset of monthly groundwater level (GWL) fluctuations was used in two observation wells (near and far from the lake) of Urmeih coastal aquifer. In the WHW model, the dataset were converted into several sub-dataset with different time scales. Then, the sub-series were used in the HW model as inputs. Subsequently, the performance of the WHW model was evaluated with ARIMA, HW and SARIMA as linear models and neural network models (ANN) and Support Vector Regression (SVR) as nonlinear models. Results showed that the WHW hybrid model could lead to a considerably increased accuracy of groundwater level monthly modeling in comparison of the linear models. The Comparison of the two wells indicated that DWT has high impact on improving the HW model performance in the well2 which located far from the lake and has less periodicity variable in the GWL series. The results also showed that the WHW model has similar accuracy in comparison with single ANN and SVR and can be helped to outperform abilities of the linear models to improve the groundwater modeling results, particularly in costal aquifers.

Keyword: Groundwater level; Holt–Winters model; Wavelet Transform; Hybrid model

1 Introduction:

Several environmental problems such as seawater intrusion, ecosystem change, and land subsidence are due to excessive groundwater pumping (Konikow and Kendy, 2005). Therefore, it is important to manage the groundwater sustainably and to plan thoroughly before the groundwater development. For effective management, Prediction of the groundwater level which has a role of fundamental indicator to estimate the quantity of groundwater has been practiced (Daliakopoulos et al., 2005; Nayak et al., 2006).

Groundwater level (GWL) can be predicted using data driven and knowledge-driven models. The performance of data driven models is based on identifying relations between input and output variables of a system without the need for experimental apparatus and complex hydro-physical models based on physical principles and mathematical equations (Solomatine et al. 2008). Required inputs of data driven models in groundwater are the hydrological and meteorological variables which are closely related to the GWL fluctuation.

The common used data driven models for simulating GWL include artificial intelligent model (e.g. ANN, ANFIS, and SVR), time series model, and combination model. The groundwater time series consists of three principal components (autoregressive, seasonality and stochastic) and the performance of the time series models are related to these components.

Based on factors influence the groundwater, time series models such as the Holt-winters (HW) model, the autoregressive moving average (ARMA) model, the Autoregressive integrated moving average model (ARIMA), and the seasonal autoregressive integrated moving average (SARIMA) model have been applied. These models are kinds of linear model, which considers linear regression or linear relationship between influence factors and the groundwater level.

The Holt-Winters method is one of the many methods or algorithms that can be used to forecast the seasonal time series. The HW model, as an Exponential smoothing method, is a univariate time series which simulates both trend and seasonal factors (Sudheer et al. 2015). The popularity of the HW model is due to the simplicity of model formulation and good forecasting results (Yang et al. 2017).

Several studies focusing on HW, ARMA, ARIMA, and SARIMA models have shown their accuracy and reliability in forecasting. Dabrowska et al. (2015) predicted the chloride content for short-term in the mineral waters of the Ustron Health Resort using SARIMA and HW models. The results of analyses indicated that the good performance of the HW model highlights its utility compared with complicated physically based numerical models. Yang et al. (2017) used three time series analysis methods, HW, integrated time series (ITS), and SARIMA to simulate the groundwater level in a coastal aquifer in China. The comparisons of three models show revealed that the HW model is more accurate in predicting the GWL than SARIMA and ITS models.

Combination different methods to constitute a hybrid model have become a popular technique for improving the precision of prediction results where the matchless specifications of all the methods are incorporated to extract various models in the

data. Empirical findings and theoretical suggestions showed that hybrid techniques could be effective in improving modeling (Kisi and Cimen 2011). Wavelet transforms (WT), as a pre-processing effective tool, can be combined with data driven models to constitute a hybrid model. Wavelet transform (WT) is a useful tool which splits up the time series into subseries containing different frequencies. The obtained subseries are very beneficial for increasing the prediction ability of a model by extracting effective information at various levels (Adamowski 2008). Recently, there has been an increasing interest in combination wavelet transform with artificial intelligence and time series models in groundwater modeling (Adamowski and Chan 2011; Moosavi et al. 2013; Maheswaran and Khosa 2013; Suryanarayana et al. 2014; Raghavendra and Deka2014; Haijiao et al. 2017).

Adamowski and Chan (2011) studied the prediction of fluctuation groundwater by a combination of wavelet transform and ANN models. They decomposed GWL and rainfall data by wavelet transform. Then, all efficient subseries of groundwater levels and rainfall were considered as inputs of the ANN model. The relative performance of the proposed coupled wavelet–neural network models (WA–ANN) was compared to regular artificial neural network (ANN) models and autoregressive integrated moving average (ARIMA) models for monthly groundwater level forecasting. The results indicated the potential of WA–ANN models in forecasting groundwater levels. Maheswaran and Khosa (2013) investigated the dynamic wavelet-based nonlinear Volterra model for long-term forecasts of groundwater. Moosavi et al. (2013) compared different data-driven techniques for predicting groundwater level for various forecasting horizons. They compared a number of model structures for ANN, ANFIS, Wavelet-ANN, and Wavelet-ANFIS to assess their ability to predict groundwater level in 1, 2, 3 and 4 months ahead under two case studies in two sub-basins. Results indicated that wavelet transform can improve the precision of groundwater level prediction. Suryanarayana et al. (2014) compared WA-SVR with ANN, SVR and autoregressive integrated moving average model (ARIMA) to show that the WA-SVR yielded better accuracy. Raghavendra and Deka (2014) applied Wavelet-Support Vector Regression (W-SVR) forecasting monthly groundwater level fluctuations observed in three shallow unconfined coastal aquifers in the southwestern part of Karnataka adjoining the Arabian Sea. To assess the accuracy efficiency of the model, The Sequential Minimal Optimization Algorithm-based SVR model was also used in the same data sets. The comparison was made with different statistical indices. Results demonstrated that WP–SVR model outperforms the classic SVR model in predicting GWL at all the three well locations. Haijiao et al. (2017) predicted GWL for lead times of 1, 2 and 3 months for 3 observation wells in the Ejina Basin using the wavelet-artificial neural network (WA-ANN) and wavelet-support vector regression (WA-SVR). Results showed that WA-ANN and WA-SVR have better performance than ANN and SVR models. WA-SVR yielded better results than WA-ANN model for 1, 2 and 3-month lead times.

It is obvious from the related literature that during the recent years, several data-driven models (e.g. ARMA, ARIMA, ANN, SVR, and hybrid models with wavelet transform) have been examined for the forecasting of groundwater. To date, no work has reported the input-output mapping capability of Wavelet-Holt Winters (WHW)

hybrid model for groundwater level forecasting. This provided an impetus for the current research. However, for the first time, this research is applied a hybrid Wavelet– Holt-Winters (WHW) technique to predict groundwater level. The main goal of this study is to examine five data driven models for the monthly forecasting of GWL in a coastal aquifer in northwest Iran, with a specific focus on new hybrid Wavelet–Holt-Winters (WHW) approach that has not been explored in the literature for groundwater level forecasting. This new hybrid model that employed wavelet transforms to decompose the time series of groundwater level into different subseries. Then each component is predicted by using the HT models and predicted components are summed to predict the original time series. Finally, the performance of Wavelet-Holt-Winters (WHW) is compared with the efficiency of single data-driven methods (i.e., ARMA, SARIMA, ANN, and SVR) which were established using the original data without de-noising processing.

2 Methodology:

1.1. Case study

The Urmeih plain is a coastal aquifer located at the east of Urmeih Lake, Iran which lies between the eastern longitude of 44°, 20′ and 45°, 20′ and northern latitude of 37°, 05′ and 38°, 05′ (Fig. 1). Urmeih Lake is a large natural reservoir, which provides water requirements for different uses such as agricultural, industrial, and domestic. The total length of coastline is around 3000 m. It is influenced under the subtropical marine monsoon climate. This plain covers an area of about 248.34 km2 and mean annual temperature and precipitation in the area are 16 °C and 304 mm, respectively. In this aquifer, groundwater flows from areas of a high hydraulic head in the west to the areas of a low hydraulic head in the east. To monitor the GWL, 55 observation wells were installed in the Urmeih aquifer. In this study, for providing a hybrid Wavelet– Holt Winters (WHW) model to predict groundwater level two observation wells were used (Well 1 and Well 2, in Fig. 1). These observation wells depicted the variable impact of the hydrogeological system in the groundwater level fluctuation at the vicinity of the lake (Well 1) and the furthest place of the lake (Well 2). Groundwater level data used in the current study cover 16 years data (192 monthly levels) between 2000 and 2016. The data set from 2000–2011 is used for model establishment, and the data set from 2012–2016 is used for predicting the dynamic change. Descriptive statistics for groundwater level at the two observation wells is shown in Table 1.

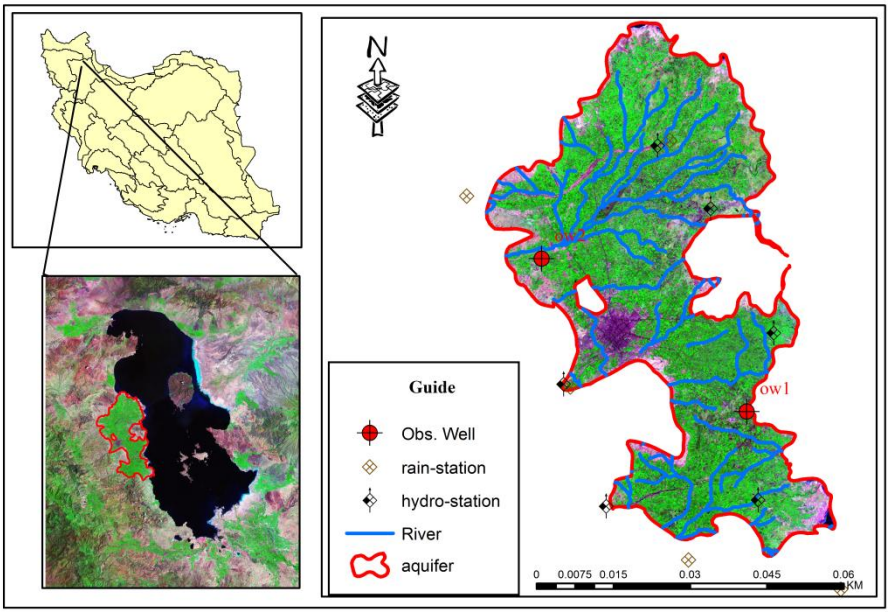


Fig. 1. The location of the study area and the spatial distributions of observation wells

Table1. Descriptive statistics for observation wells.

Obs. Well	Statistic Parameters (m)			
	Mean	*Max*	*Min*	*StD*
OW1	**1277.01**	**1279.37**	**1274.91**	**1.11**
OW2	**1343.22**	**1362.41**	**1323.29**	**7.18**

1.2. Holt–Winters model

Exponential smoothing is a pragmatic approach to forecasting whereby the prediction is constructed from an exponentially weighted average of past observation. The Holt–Winters model, as an exponential smoothing method, is the one most commonly used (Sudheer et al.2014). The basic structures were provided by Holt and Winters (Holt 1957). Its basic idea is to decompose a time series into a linear trend component, seasonal variation component, and random change component, incorporating the exponential smoothing algorithm; the long-term trend (S_t), trends incremental (b_t), and seasonal changes (I_t) are estimated; and then, a predictive model is established to extrapolate the predicted value. It is an extended single exponential smoothing, which allows forecasting data with non-constant trends and seasonal variations. Thus, it is also capable for detecting trend in different time periods. That the Holt–Winters method is capable in tracking changes in the level, trend, and also seasonality; the influence of the random motion can be moderately filtered Kamruzzaman et al.

(2011). Therefore, it is particularly suitable for time series prediction containing the trend and seasonal variation. The Holt–Winters exponential smoothing approach includes methods for both additive and multiplicative seasonal patterns. The Holt–Winters method works with the following recursive updating equations:

$$F_t = \alpha(F_{t-1} - T_{t-1}) + (1 - \alpha)\frac{Y_{t-1}}{S_{t-K}} \quad (1)$$

$$S_t = \Delta S_{t-K} + (1 - \Delta)\frac{Y_t}{F_t} \quad (2)$$

$$T_t = \gamma T_{t-1} + (1 - \gamma)(F_t - F_{t-1}) \quad (3)$$

$$F_{t+k} = F_t + kT_n + S_{t+k-c} \quad (4)$$

Here α , Δ and γ are the smoothing parameters associated with level, trend and seasonality, respectively, which are estimated by minimizing the sum of squares of errors or any suitable error function. These coefficient ranges between 0 and 1. While c is the length of the seasonal cycle, k is the time intervals from the moment to the predicted time, and F_{t+k} is the predicted value in HW model.

1.3. Wavelet Transform

The wavelet transform is the popular tool of the Fourier Transform because of its multi resolution in time and frequency domain. The function ψt presents a mother wavelet, which has shock characteristics and diminishes to zero quickly (Zhou et al. 2008). Wavelet functions decompose the time series into subseries with various frequency domains, and then considers each subseries with a resolution corresponding to its frequencies thus dominating the deficiencies of Fourier transform. In employing Discrete Wavelet Transform (DWT), a finite number of shifts and scale levels are considered. The wavelet coefficients are calculated using DWT and the simplest and most impressive approach, where scales (a) and shift (τ) are chosen on basis of the powers of 2, called dyadic scales and shifts (Mallat 1989). DWT functions are usually presented by

$$a = a_0^j \quad \tau = ka_0^j\tau_0 \qquad a_0 > 0,$$
$$\tau_0 \in R, \ \forall j,k = 0,1,2,3,\ldots, \ m \in Z \quad (5)$$

$$\psi_{j,k}(t) = a_0^{-j/2}\psi\left(\frac{t - ka_0^j\tau_0}{a_0^j}\right) = a_0^{-j/2}\psi\left(ta_0^{-j} - k\tau_0\right) \quad (6)$$

The most common selection is a0 =2, τ0 =1, and then the DWT becomes binary. For a discrete lake level time series LL(t), in which LL(t) happens at a discrete integer time step t, the dyadic discrete wavelet transformation can be defined as:

$$W_f(j,k) = \sum_{i,k \in z} LL(t)2^{-j/2}\psi\left(2^{-j}t - k\right) \quad (7)$$

Here, Wf (j,k) shows the specifications of the lake level time series in scale (a or j) and time domain (τ or k). When either a or j is small, the frequency resolution diminishes, but the time domain increases. When either a or j is large, the frequency resolution increases, but the time domain diminishes (Wang and Ding 2003). The input signal can be built using the equation:

$$LL(t) = \sum_{j,k\in} W_f(j,k)\psi_{j,k}(t) \quad (8)$$

In Eq. (8), Wf (j,k) applies down-sampling to calculate an approximation coefficient (aI) at decomposition level I with a low pass filter $L(\psi_{i,k}(t))$, and detail coefficients (d1, d2, d3, . . ., dI) at various levels 1, 2, . . ., I with a high pass filter $H(\psi_{i,k}(t))$. aI presents identity of the time series and d1, d2, d3, . . ., dI show the detailed information of the time series such as jump, period, break and so on. However, the coefficient aI and dI cannot be directly summed to generate the time series as they are produced by down-sampling and are only half the length of the time series. So, it is indispensable to reconstruct the approximations (aI) and details (dI) before summing them (Wei et al. 2013). Then the time series can be explained as:

$$LL\ (t) = a_I L\left(\psi_{j,k}(t)\right)\ + \sum_{I=1} d_I H\left(\psi_{j,k}(t)\right) \quad (9)$$

$$LL\ (t) = a_I\ +\ \sum_{I=1} d_I \quad (10)$$

1.4. Wavelet-Holt Winters (WHW) hybrid model

The WHW model utilizes the wavelet transform and HW model to propose a hybrid method for Exponential smoothing. This hybrid method is based on the idea that discrete wavelet transform, as a multiscale decomposition tool, is applied to decompose the original data into several other scales (sub series) with different levels of resolution. In fact, the DWT splits up the original data into one approximation series (A) and N detail series in the wavelet domain (D1, D2, …, D_N) Where N implies to decomposition levels. To determine the decomposition levels, Wang and Ding (2003) suggest a formula (I=int[log(N)], where I is length of time series data used in training) for the maximum decomposition level. The historical groundwater level data (H) obtain using the relation $H = \mathrm{A} + \sum_{n=}^{N} \mathrm{D}_N$. The best resolution level (N) is chosen based on the approximation signal's ability to describe the general pattern of the load series. These sub series are very beneficial for increasing the prediction ability of a model by extracting effective information at various levels. Then, each sub- series is considered as inputs of the HW model to forecast. For HW model, assigning the appropriate smooth coefficients α, Δ, and γ is important. It should be noticed that Due to the influence of various factors, uncertainty and randomness exist in groundwater table depth. Hence, the different selection of smooth coefficients still inevitably leads to the unreliability of prediction. In this study, the optimal value of smooth coefficients was specified based on a trial and error approach. As the minimum value of the mean squared forecast error determines the optimum values of the parameters. Finally, results of sub series forecasting are Summarized and generated forecasted groundwater level ($F_H = F_{AH} + F_{D1} + F_{D2} + F_{D3}$). A schematic structure of the WHW hybrid model was shown in fig.2 .To establish DWT and HW modeling were done using MATLAB and MINITAB software, respectively.

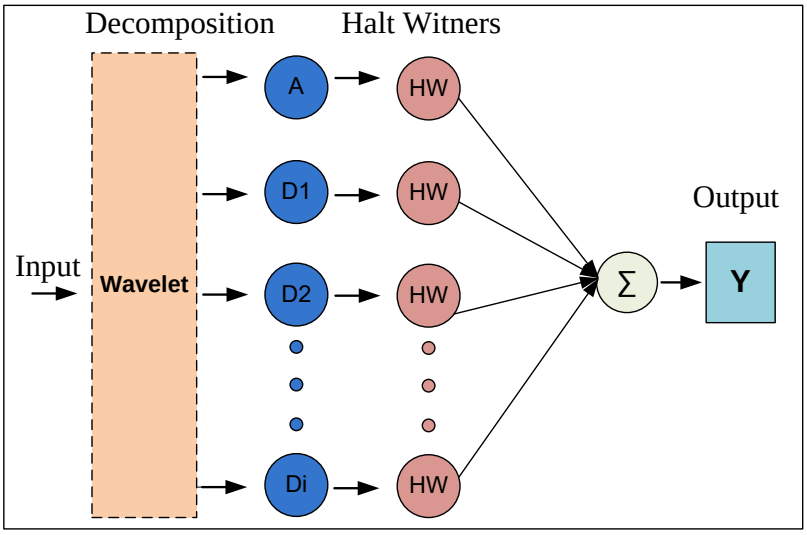


Fig.2. schematic structure of the WHW hybrid model

3 Results:

In this paper, the new WHW hybrid model was applied for the first time to forecast monthly GWL for 2 observation wells in the Urmeih coastal aquifer. The performance of this hybrid model was compared with the efficiency of single data driven methods (i.e., ARMA, HW, SARIMA, ANN, and SVR) which were established using the original data without de-noising processing. In the first step, the historical groundwater level data (H) was decomposed via the DWT into some levels of decomposition. When performing DWT, the wavelet mother and decomposition levels must be carefully selected (Haijiao et al. 2017). Three decomposition levels were applied by the suggestion of Wang and Ding (2003) in order to decomposition of time series. The monthly groundwater level data were decomposed using Daubechies (db4) mother wavelet. DWT breaks up original time series into four subseries containing three detail components and one approximation component. Each component plays a particular role in time series and follows a special pattern about the original time series. For this reason, all the components were applied as inputs to the HW model. It is important in model development that the full data split to training and test data as a requirement to avoid incorporating future information (Haijiao et al. 2017). Hence, the wavelet decomposition process was applied on each partition (i.e., training and testing) independently. It should be emphasized that if the full dataset (i.e., training and testing) is decomposed, then future data (that is not available to the modeler) could be used in the calculation of wavelet coefficients and lead to introduce bias into the forecasts (Deo et al. 2016).

In the second step, various input combinations of subseries were tried for predicting subseries, separately. The multiplicative method is selected in MINITAB software to establish the HW model and the optimal value of smooth coefficients was specified

based on a trial and error approach. Selection of optimal smooth coefficients and best prediction model are determined by the minimum of root mean square error (RMSE). After calculation, the RMSE was found to get the minimum; therefore, for each subseries, a combination of smoothing coefficient was considered as the optimal prediction model. The optimal smooth coefficients and fitting curve between observed subseries and the calculated subseries and for Well1 and Well2 were shown in Fig. 3 and 4, respectively. Finally, the groundwater level was predicted by summation of each subseries' forecasts obtained using the optimal HW models.

The accuracy of the WHW hybrid models was evaluated with the performance of single data driven models (i.e., ARIMA, HW and SARIMA as linear models, ANN and SVR as nonlinear models) which were established using the original data without de-noising processing. In nonlinear models, the time series of groundwater level, Lake water level, precipitation, and hydrometric data were used as input data. In ARIMA and SARIMA models, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the groundwater data were extracted. Then, parameters (p,d,q, P, D, Q) was defined considering the cutout and continuity features in the plot of ACF and PACF. In the ANN model, the weights were adapted using LM learning algorithm, the tansigmoid transfer function was used for the hidden nodes and the size of the hidden nodes was identified using a trial and error approach. The radial basis function (RBF) was used as the kernel function of the SVR model and it optimal parameters were determined based on a trial and error approach. The R, RMSE and NSE statistics of the optimal WHW model and single data driven models are given in Tables 2 for Well1 and Well2.

With respect to R, RMSE and NSE criteria, The WHW model has better performance than the HW model which shows the wavelet transformation abilities in the multiscale decomposition of time series and also HW model capabilities in simulating each subseries. It was concluded that the HW model with an autocorrelation structure cannot simulate a multiple seasonality time series such as coastal groundwater level fluctuation appropriately. In fact, DWT by separating of non-linear, long-time, and random terms in groundwater could be helpful to transform non-linear and non-stationary time series to independent subseries which is useful to improve the HW model accuracy.

Comparison of the two wells indicates that DWT has had a high impact on improving the HW model performance in the well OW2, so that the RMSE and NSE criteria in the optimal WHW in compare with WH model improve 66% and 30%, respectively. While, in Well1 the RMSE and NSE criteria improve 14% and 8%, respectively. The main reason for this may be the fact that more periodicity exists in the groundwater level series at Well1, which located near the lake, in compare with Well2. The existence of more periodicity in Well1 leads to the good performance by original HW model. This result reflects that if there are multiple seasonal fluctuations and periodicity in the groundwater time series, then the performance of the WHW model will be better. It should be noticed that the more periodicity is considerably affected by the seasonal fluctuations of Urmeih lake due to the direct influence of the natural factor of climate on the lake (Nourani et al. 2016).

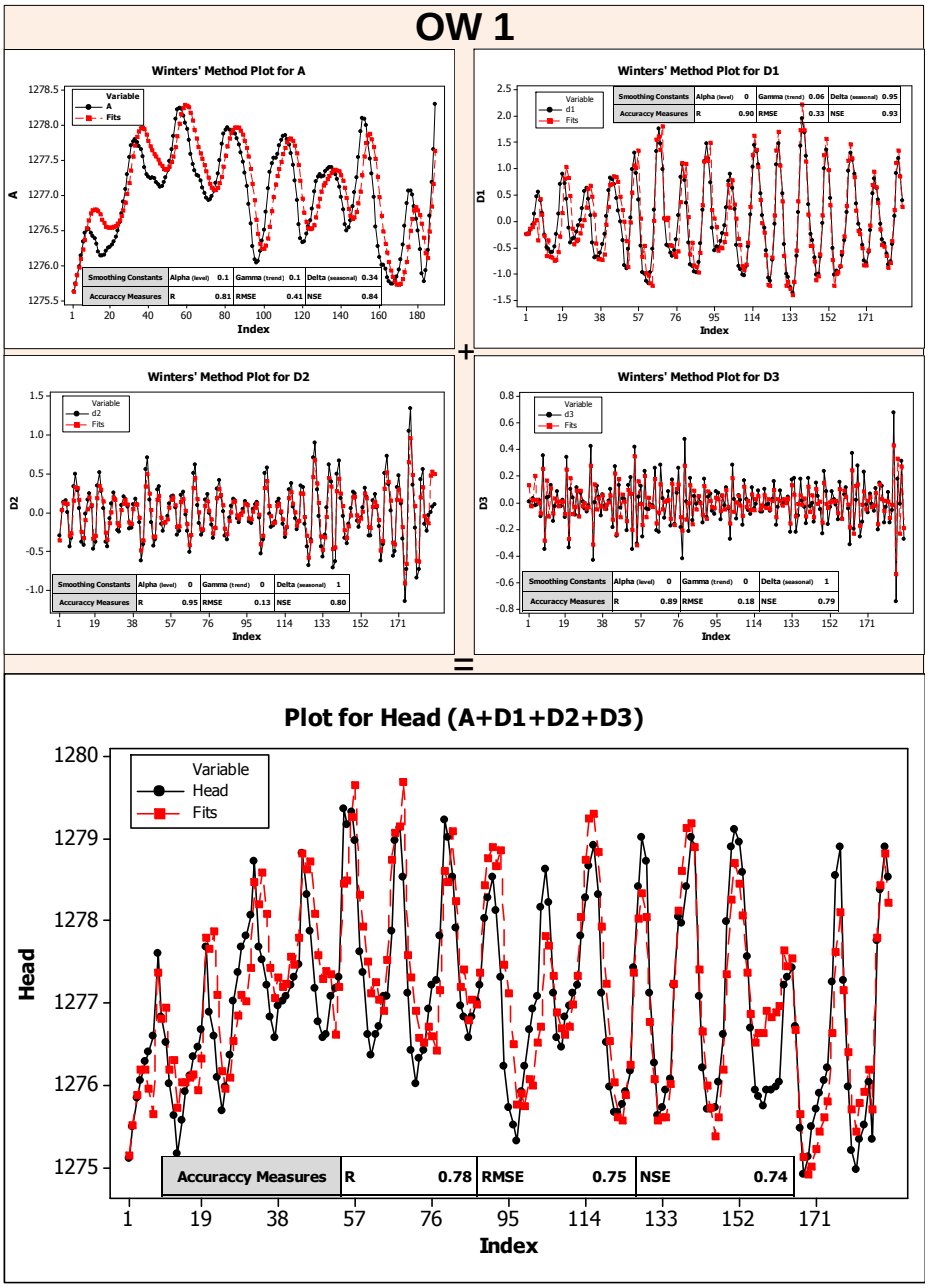


Fig.3. The optimal smooth coefficients and fitting curve between observed subseries and the calculated subseries at Well1

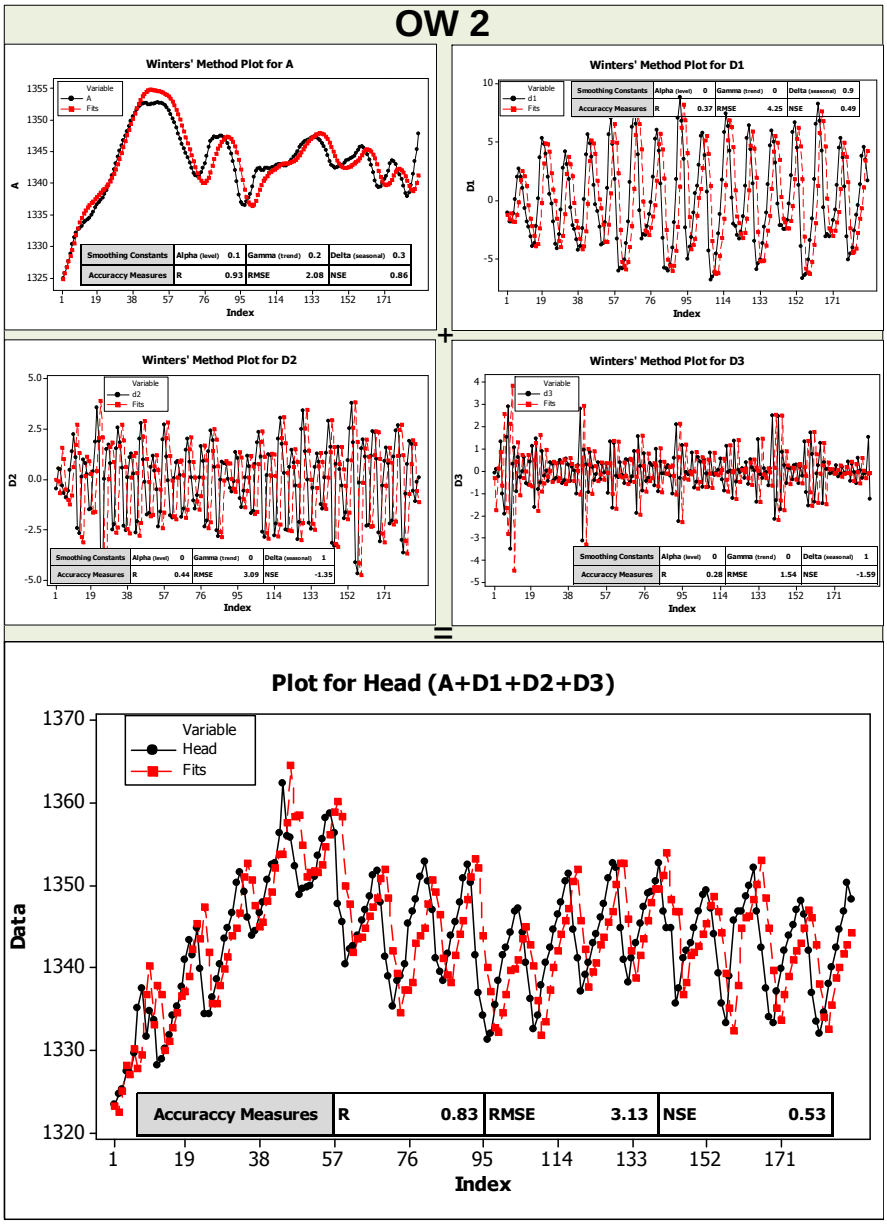


Fig.3. The optimal smooth coefficients and fitting curve between observed subseries and the calculated subseries at Well2

It can be observed from the statistical results in Table 2 that the WHW model was found to outperform ARIMA, SARIMA, and WH models. WHW model has higher R and smaller RMSE than ARIMA, SARIMA, and WH models in both wells.

The results of in Table 2 also shows that the nonlinear models, ANN and SVR models, provide slightly better performance than the WHW model due to using a suitable algorithm for prediction (Figs. 4 and 5). It can be concluded that a WHW model has similar accuracy in comparison with single ANN and SVR and can be helped to outperform abilities of the linear models to improve the groundwater modeling results, particularly in coastal aquifers.

The All models in Well 1 consistently showed higher R and lower RMSE values compared to the Well 2. It is observed a worth note that there are more stochastic and unknown parameters in well 2 and data driven models are not able properly to predict this time series.

Table 2. Performance of WHW, linear and nonlinear models for groundwater level forecasting at Well 1 and Well 2

Obs. well	Data-Driven models	Structure	Trianing Step			Testing Step		
			R	RMSE	NSE	R	RMSE	NSE
OW1	WHW	db4 , i=3	0.87	0.67	0.67	0.78	0.75	0.74
	HW(α,γ,β)	(0.2,0.1,0.3)	0.81	0.64	0.79	0.69	0.86	0.68
	ARIMA(p,d,q)	(1,1,1)	0.67	1.28	0.61	0.72	1.60	0.55
	SARIMA(p,d,q)(P,D,Q)	(1,1,1)(2,1,1)12	0.73	0.82	0.66	0.80	1.20	0.63
	ANN	(4,3,1)	0.85	0.54	0.72	0.86	0.67	0.73
	SVR(γ, ε, C)	(10,0.1,15)	0.90	0.43	0.82	0.90	0.59	0.78
OW2	WHW	db4 , i=3	0.88	1.12	0.73	0.83	3.13	0.53
	HW(α,γ,β)	(0.2,0.1,0.3)	0.94	1.44	0.89	0.78	5.20	0.45
	ARIMA(p,d,q)	(1,0,1)	0.84	2.77	0.77	0.75	6.20	0.34
	SARIMA(p,d,q)(P,D,Q)	(1,0,1)(2,1,1)12	0.96	1.89	0.94	0.69	5.77	0.63
	ANN	(4,2,1)	0.93	2.81	0.86	0.88	2.71	0.77
	SVR(γ, ε, C)	(1,0.1,20)	0.92	3.01	0.84	0.85	3.28	0.63

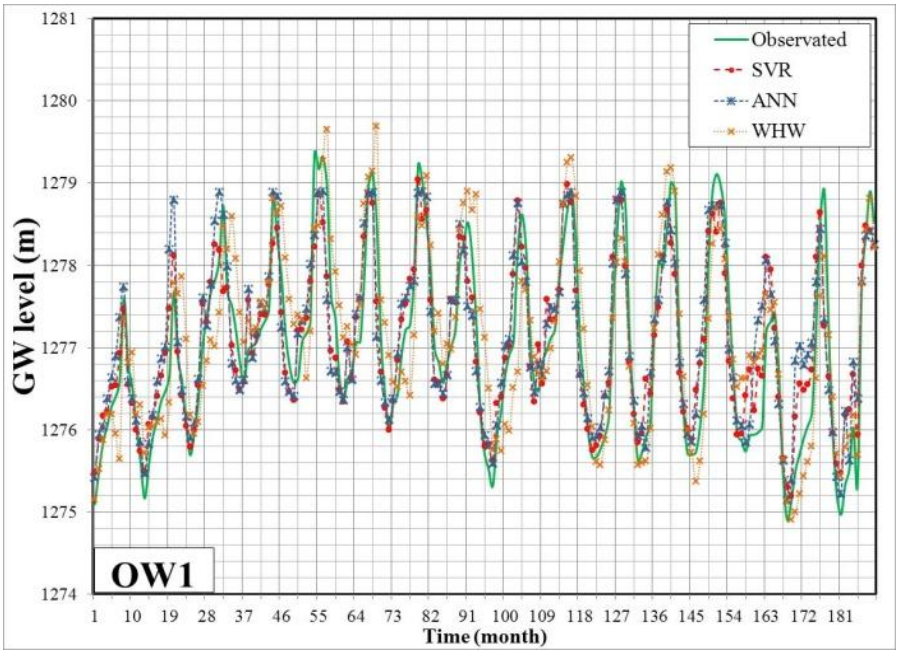


Fig.4. Results of training and testing steps for the WHW, ANN, and SVR models at Well 1

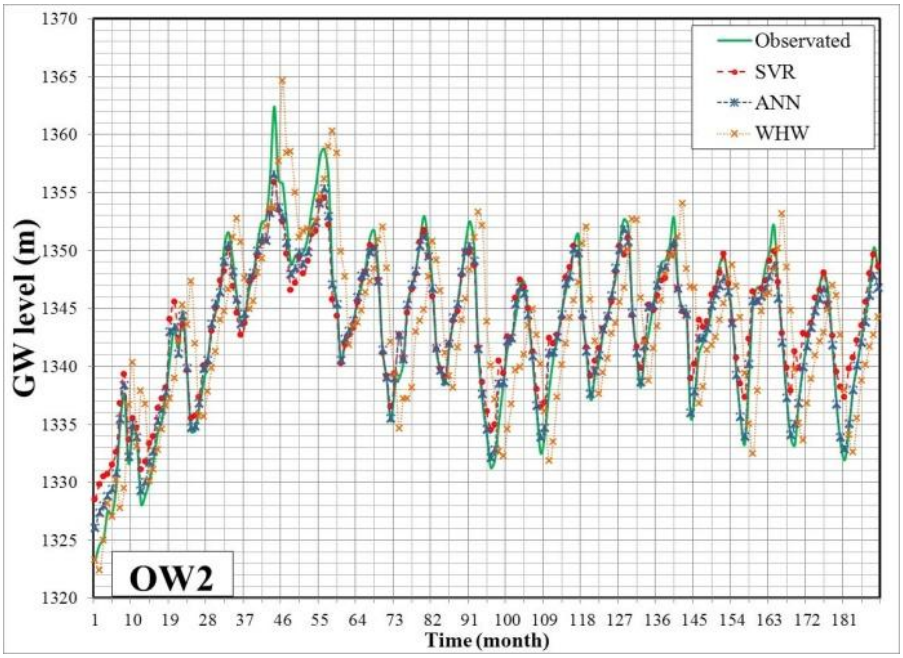


Fig.5. Results of training and testing steps for the WHW, ANN, and SVR models at Well 2

4 Conclusions

The Holt–Winters (WH) model, as an exponential smoothing method, is capable to predict the time series containing the trend and seasonal variation and is a more sophisticated method of forecasting than of moving average methods (Young et al. 2017). This model has not an appropriate performance in multi-scale and multi-frequency series time. In order to the solution this problem and improve the HW model, a new method based on the conjunction of discrete wavelet transform (DWT) and HW model was proposed. In the groundwater time series generally there are multiple seasonal fluctuations that this hybrid model proposed can be used to as an effective method for simulation. Unlike the HW model that directly uses the original data as input data, the WHW hybrid model use an approximate part associated with a low frequency and some detailed part associated with high frequencies as input data that was decomposed by the wavelet transform. In this study the WHW hybrid model was applied for the first time in GWL modeling. Also, a comparative study was performed to investigate the feasibility of WHW model to simulate GWL at two observation wells (near and far from the lake) in the Urmeih coastal aquifer in northwest, Iran. The monitored observation monthly GWL data from 2000 to 2016 were used in model setup and prediction. The monthly GWL data were decomposed using Daubechies (db4) mother wavelet with three decomposition level. Finally, the WHW model was compared to linear and nonlinear models (i.e., HW, ARMA, SARIMA, ANN, and SVR) which were established using the original data without de-noising processing. In nonlinear models, the time series of groundwater level, Lake water level, precipitation, and hydrometric data were used as input data. Results showed that the DWT could lead to a considerably increased accuracy of GWL modeling by HW model. However, the WHW hybrid model is more accurate in predicting the GWL than HW, ARIMA, and SARIMA models. The WHW hybrid model also has the same performance compared to nonlinear models. By comparison two wells, this research reflects that if there are multiple seasonal fluctuations in the groundwater time series, the performance of the WHW model will be better.

5 Refference.

1. Adamowski, J. and Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology, 407(1-4), pp.28-40.
2. Adamowski, J.F., 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. Journal of Hydrology, 353(3-4), pp.247-266.
3. Daliakopoulos, I.N., Coulibaly, P., and Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural network, J. Hydrol. Vol. 309, pp.229–240, 2005.

4. Deo, R.C., Tiwari, M.K., Adamowski, J.F. and Quilty, J.M., 2017. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. Stochastic environmental research and risk assessment, 31(5), pp.1211-1240.
5. Djurovic, N., Domazet, M., Stricevic, R., Pocuca, V., Spalevic, V., Pivic, R., Gregoric, E. and Domazet, U., 2015. Comparison of groundwater level models based on artificial neural networks and ANFIS. The Scientific World Journal, 2015.
6. Guzman, S.M., Paz, J.O., Tagert, M.L.M. and Mercer, A., 2015. Artificial Neural Networks and Support Vector Machines: Contrast Study for Groundwater Level Prediction. In 2015 ASABE Annual International Meeting (p. 1). American Society of Agricultural and Biological Engineers.
7. Holt, C.C., 2004. Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting, 20(1), pp.5-10.
8. Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence. Prentice-Hall, Eaglewood cliffs. doi:10.1109/tac.1997.633847
9. Kisi, O, and Cimen. M., 2009. Evapotranspiration modelling using support vector machines. Hydrological sciences journal 54.5 (2009): 918-928.
10. Kisi, O., Shiri, J., 2012. Wavelet and neuro-fuzzy conjunction model for predicting water table depth fluctuations. Hydrology Research 43 (3), 286–300.
11. Mayilvaganan MK, Naidu KB (2010) Comparative Study of ANN and ANFIS for the Prediction of GWL of a Watershed. Global J Math Sci: Theory and Practical 3: 299-306.
12. Mukherjee, A. and Ramachandran, P., 2018. Prediction of GWL with the help of GRACE TWS for unevenly spaced time series data in India: Analysis of comparative performances of SVR, ANN and LRM. Journal of Hydrology, 558, pp.647-658.
13. Nayak, P.C., Rao, Y.S. and Sudheer, K.P., 2006. Groundwater level forecasting in a shallow aquifer using artificial neural network approach. Water Resources Management, 20(1), pp.77-90.
14. Shiri, J., Kisi, O., Yoon, H., Lee, K.K., Nazemi, A.H., 2013a. Predicting groundwater level fluctuations with meteorological effect implications: a comparative study among soft computing techniques. Comput. Geosci. 56, 32–44.
15. Shirmohammadi, B.; Vafakhah, M.; Moosavi, V.; Moghaddamnia, A. Application of several data-driven techniques for predicting groundwater level. Water Resour. Manag. 2013, 27, 419–432
16. Solomatine, D.P., Maskey, M. and Shrestha, D.L., 2008. Instance- based learning compared to other data- driven methods in hydrological forecasting. Hydrological Processes: An International Journal, 22(2), pp.275-287.
17. Sujay Raghavendra, N. and Deka, P.C., 2015. Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression. Cogent Engineering, 2(1), p.999414.

18. Suryanarayana, C., Sudheer, C., Mahammood, V. and Panigrahi, B.K., 2014. An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India. Neurocomputing, 145, pp.324-335.
19. Wang, J., Wang, J., Li, Y., Zhu, S. and Zhao, J., 2014. Techniques of applying wavelet de-noising into a combined model for short-term load forecasting. International Journal of Electrical Power & Energy Systems, 62, pp.816-824.
20. Wang, W. and Ding, J., 2003. Wavelet network model and its application to the prediction of hydrology. Nature and Science, 1(1), pp.67-71.
21. Yu, H., Wen, X., Feng, Q., Deo, R.C., Si, J. and Wu, M., 2018. Comparative study of hybrid-wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, Northwest China. Water Resources Management, 32(1), pp.301-323.

Tipping point analysis and its applications in geophysics, environmental sciences, and smart sensor systems

V.N.Livina

National Physical Laboratory, Teddington, Middlesex, United Kingdom
valerie.livina@npl.co.uk

Tipping point analysis techniques developed in the past few years help anticipate, detect and forecast tipping points in a dynamical system. The methodology combines monitoring short- and long-term memory in a time series with potential analysis that analyses and extrapolates the system states.

For anticipating tipping points, early warning signal (EWS) indicators include dynamically derived lag-1 autocorrelation (ACF), power-law scaling exponent of Detrended Fluctuation Analysis (DFA, [2]), and recently developed power-spectrum-based EWS indicator, which similarly monitors dynamical changes memory in a record [3]. When such values rise monotonically, this indicates an upcoming transition or bifurcation in a series and can be used as early warning signals for tipping.

The basis of the early warning signals is that the system potential $U(t)$ described in general form by a polynomial equation (see [4]) of the state equation

$$\dot{z}(t) = -U'(z) + \sigma\eta$$

start experiencing shallowing of an existing or growing of a new potential well. This causes the decay rate of the system fluctuation to slow down, which is often accompanied by increasing memory in the time series. This can be detected and quantified by various scaling analysis techniques, such as DFA [2] or power spectrum [3]. Even the simple lag-1 autocorrelation function can be used to detect this effect [1], however it may be affected by trends that autocorrelation function cannot account for, which is a shortcoming of the ACF-indicator.

By combining several EWS indicators, it is possible to identify and distinguish different types of tipping, such as forced transitions (when the underlying trend increases autocorrelations, but the detrended fluctuations remain unchanged) and genuine bifurcations (when the structure of the system potential changes, i.e. a new potential well appears, or an existing well disappears). Quantification of the scaling properties of the fluctuations also indicates the proximity of the tipping point in gradually varying systems; in the case of an abrupt tipping point, however, EWS techniques are often capable only to detect the change rather than anticipate it, due to the high speed of the system dynamics, with a very short forewarning time interval.

The potential analysis detects a transition or bifurcation in a series at the time when it happens, which is illustrated in a special colour plot mapping the potential dynamics of the system [4-8]. This technique provides a 2D "portrait" of a time series with multiscale mapping of the number of detected system states marked by different colours,

so that a system bifurcation is visible, similar to how variable periodicities are illustrated by wavelet analysis plots.

Potential analysis is also used in forecasting time series by extrapolation of Chebyshev approximation coefficients of the kernel distribution, with reconstruction of correlations in the data. First, coefficients of the Chebyshev approximation of probability density are collected in sliding windows; second, the series of these coefficients are extrapolated and used to reconstruct the forecast probability density; third, a time series is simulated from the obtained probability density using rejection sampling; fourth, the time series is sorted according to historic data, with reconstruction of temporal correlations, taking into account seasonality if necessary (this is important in geophysical and environmental series, where there is seasonal trend). For further details of the algorithm, see [8].

The methodology has been extensively tested on artificial data and on various observed datasets [2-14], and proved to be applicable to trajectories of dynamical systems of arbitrary origin [15]. We will demonstrate the methodology on diverse examples of synthetic and observed time series, in particular, from climatology [3,4,7,14], ecology [11], structural health monitoring [10,13], and acoustics [14].

References:

1. Held and Kleinen, GRL 2004;
2. Livina and Lenton, GRL 2007;
3. Prettyman et al, EPL 2018;
4. Livina et al, Climate of the Past 2010;
5. Livina et al, Climate Dynamics 2011;
6. Livina et al, Physica A 2012;
7. Livina and Lenton, Cryosphere 2012;
8. Livina et al, Physica A 2013;
9. Drijfhout et al, PNAS 2013;
10. Livina et al, Journal of Civil Structural Health Monitoring, 2014;
11. Kefi et al, PLoS ONE 2014;
12. Livina et al, Chaos 2015;
13. Perry et al, Smart Materials and Structures, 2016;
14. Livina et al, NPG 2018;
15. Vaz Martins et al, PRE 2010.

Combination of neural network and wavelet to predict suspended sediment load in river by using data clustering

Samir Bengherifa[1], Abdelouahab Lefkir[2] and Abdelmalek Bermad[1]

[1] ENP, Polytechnic School of Algiers, Material and Environment Laboratory, Bp 32 Hassen-Badi 16000, Algeria
bengherifa07@yahoo.fr
[2] ENSTP, Ecole Nationale Supérieure des Travaux Publics, LTPiTE Laboratory, Algeria

Abstract. Predicting suspended sediment load in river is an important task in order to manage and solve water resources problems. Neural network models have been developed to predict daily suspended sediment load (Q_s) for Eel river at Scotia, California, USA. A 21 years daily time series data were used in the present study, the time series of flow load (Q_l) and suspended sediment load (Q_s) are affected by seasonal irregularity and non-stationarity flow. The present study analyzes the effect of data clustering and wavelet transform on neural network models. The results showed that wavelet transform combined with neural network (WANN) with one cluster has higher performance than artificial neural network with either one cluster or three clusters.

Keywords: Artificial Neural Network, Wavelet, Suspended Sediment Load, Clustering.

1 Introduction

The prediction of the volume of sediment carried out by river is required for planning and management of water resources projects. It's an important task in terms of rivers navigability, reservoir capacity, pollution and water quality, and life of hydroelectric equipment [Ozgur Kisi et al. 2012].

Several studies in recent decades have been conducted to estimate the daily suspended load Q_s using a combination of Wavelet and Artificial Neural Network (WANN), Taher Rajaee [2011] applied two models ANN and WANN to estimate the suspended sediments load and found that WANN has superior performance than ANN. Daubechies Wavelets have been used by many authors for its effectiveness in modeling hydrology time series [Vahid Nourani et al. 2015].

The present study investigates two models, ANN and WANN, applied to the time series of daily flow discharge Q_l and Q_s by using one cluster and three clusters. The time series of the Q_l and Q_s were arranged only in one cluster in the first case, and in the second case the data were arranged in three clusters from low values, medium values, and high values of the times series Q_l and Q_s. Multi-layer perceptron

(MLPs) is used for ANN, Daubechies wavelet is used to filter the input signal discharge Q_l.

2 Methods or Materials

2.1 wavelet analysis

The wavelet transform of a signal is capable of providing time and frequency information simultaneously. In order to complete this task, the time series are broken down by the transformation into wavelets that are scaled and shifted version of the mother wavelet (Nason et al. 1999). When applying wavelet transform, the choice of the mother wavelet is a crucial task. For the Q_l and Q_s time series in this study, Daubechies wavelet of order 8 (Db8) has been applied following recommendations given by Rajaee et al. (2009). An automatic denoising was performed to eliminate any existing noise in the signal Q_l, by using soft Donoho and Johnstone's universal threshold with level-dependent estimation of the noise. After denoising the flow discharge Q_l, the time series will be subject to ANN model for prediction.

2.2 Artificial neural network

The feed-forward multi-layer perceptron is the most popular of ANN models for its wide applications, which usually uses the technique of error back-propagation to train the network (Ehsan O. et al. 2015).The neural network used in this study has three layers, an input layer, a hidden layer, and an output layer. For the network training, the Bayesian Regularization back-propagation algorithm was adopted (Raveendra K et al. 2008). Also, the transfer functions consist of Elliot sigmoid function for the hidden layer and linear function for the output layer. To compute the number of neuron in the hidden layer of the ANN model, the common trial and error method was adopted (Qian-Jin Linet al. 2013). The ANN network was trained up to thirty neurons to optimize its performance. The number of neurons was not kept the same for ANN and WANN models. The input layer has only one input which the flow discharge Q_l at the time current t, the output layer has only one output which is Q_s at the time t.

2.3 Gauging station and model performance evaluation

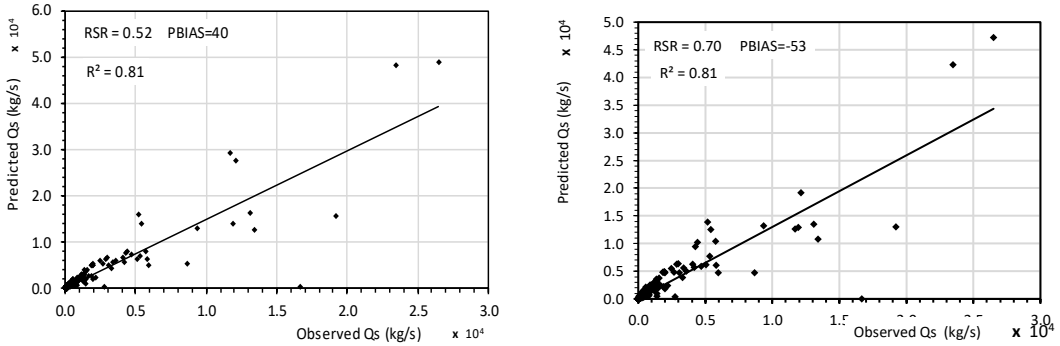
The river discharge Q_l and the suspended sediment load Qs data from the Eel river at Scotia, California, USA (USGS station N° 11477000, basin area of approximately 3600 square miles, Latitude 40°29'30", Longitude 124°05'55"), were employed to train and test all the models developed in this study. The time series data are from October 1959 to September 1980, which represent a total of

7670 days. The data were divided in three parts, training data, validation data, and testing data.

Model The performance of the model is based on two criteria, the goodness-of-fit, where R^2 and NSE were used, and the error, where RSR and PBIAS were used.

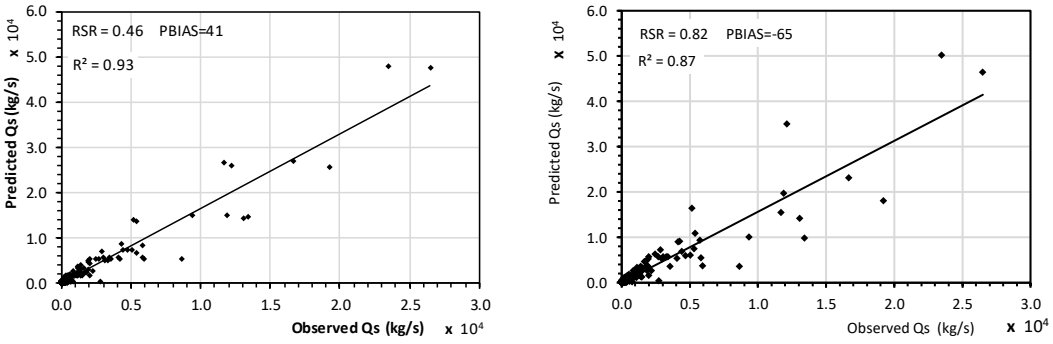
3 Results

Two models ANN and WANN were employed with one cluster and three clusters, to model the relationship between the flow discharge Q_l and the suspended sediment discharge Q_s. Fig.1 and Fig 2 display, respectively, ANN and WANN models.



One cluster Three clusters

Fig. 1 Observed and simulated Q_s of the ANN model for the test phase



One cluster Three clusters

Fig. 2 Observed and simulated Q_s of the WANN model for the test phase

All statistical analyses are presented in table 1. The bold numbers relatively to high values of R^2 and NSE and relatively to low values for RSR and absolute value of PBIAS

Table 1 Statistical analysis of ANN and WANN models by using one cluster and three clusters

Models	ANN	WANN	ANN	WANN
Clustering	1 cluster	1 cluster	3 clusters	3 clusters
Number of neurons in hidden layer	27	29	21-27-2	3-30-24
R^2	0.81	**0.93**	0.81	0.87
RSR	0.52	**0.46**	0.70	0.82
NSE	0.72	**0.79**	0.51	0.32
PBIAS	**40**	41	-53	-65

4 Discussion

Both ANN with one cluster model and ANN with 3 clusters have the same value of coefficient of determination (R^2 =0.81); in contrast, Nash-Sutcliffe efficiency coefficient had net higher value (NSE=0.72) for ANN with one cluster than ANN with 3 clusters model (NSE=0.51), which implies that the fit line of ANN-1 cluster model estimate is closer to the observed data than the one from ANN-3 clusters model. Also, ANN-1 cluster model had less error, with lower values of RSR and PBIAS than ANN-3 clusters model.

The aim of the wavelet is to improve the performance of the ANN model. Table 1 revealed that WANN-1 cluster model, over all, has superior performance (R^2=0.93, RSR=0.46, NSE=0.79, PBIAS=41) than ANN-3clusters (R^2=0.87, RSR=0.82, NSE=0.32, PBIAS=-65). Therefore, WANN with one cluster estimates the sediment load better than ANN. In contrast, WANN with 3 clusters model performed poorly against ANN with 3 clusters.

5 Conclusions

In this study, two different neural network models (ANN and WANN) were trained to predict sediments load in Eel river, located in Scotia, California (USA). The neural network input was processed through discrete wavelet transforms and data clustering, to take in consideration the non-stationarity and the seasonal irregularity in the flow discharge and sediment load time series. Results show that the wavelet transform improved the estimation of the sediment load when using one cluster for the time series.

References

1. Barbara Cannas, Alessandra Fanni, Linda See, Giuliana Sias. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. Physics and Chemistry of the Earth, *31*(18), 1164–1171(2006).
2. Nason, G.P., Von Sachs, R., 1999. Wavelets in time series analysis. Philosophical Transactions of the Royal Society of London Series A 357, 2511–2526.
3. Nourani V, Andalib G Daily and monthly suspended sediment load predictions using wavelet b ased artificial intelligence approaches. Journal of Mountain Science 12(1) (2015).
4. Ozgur Kisi, Jalal Shiri. River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques. Computers & Geosciences 43:73–82(2012).
5. Ozgur Kisi. Wavelet Regression Model as an Alternative to Neural Networks for River Stage Forecasting. Water Resour Manage 25:579–600(2011).
6. Qian-Jin Liu, Zhi-Hua Sh, Nu-Fang Fang, Hua-De Zhu, Lei Ai. Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet–ANN approach. Geomorphology 186: 181–190(2013)
7. Rajaee, T.; Mirbagheri, S. A.; Nourani, V.; Alikhani, A. Prediction of daily suspended sediment load using wavelet and neurofuzzy combined model. Int. J. Environ. Sci. Tech., 7 (1), 93-110(2009).

Investigation and Forecasting of Hydrological Time Series

Dr. Sc. Ing. Svetlana Polukoshko

Ventspils University College, Engineering Research Institute "VSRC". Inzenieru Str. 101, Ventspils, LV-3601, Latvia.

Email: pol.svet@inbox.lv

Abstract. In this paper the average monthly river discharge and mean annual sea level in the coastal zones are discussed because of the great importance of this indicators in the production activity of people. These indicators form non-stationary nonlinear time series, therefore their investigation are performed with help of singular spectrum analysis (SSA) technique. SSA is powerful model-free method for time series analysis and forecasting. It allows to decompose the non-stationary time series into trend (slowly varying components), periodic or seasonal components (sum of elementary harmonic), noise (random components) and to forecast subsequent behavior of system, if separability condition of time series and possibility of reverse synthesis of individual elements are fulfilled.

Keywords: Nonstationary time series, Mean sea level, River discharge, Singular spectrum analysis, Trend, Periodical components, Forecasting

1 Introduction

As soon as the water is the primary factor of the survival of mankind, water objects observations are performed and data are stored from old times, but the analysis of this supervision results began only in recent years. For this purpose various methods may be used, among them model–free SSA technique of time series analysis is very convenient tool because of the possibility of trend, seasonal periodical components and noise extraction and also forecasting series behaviour and all its components separately [3], [4], [5]. Any hydrological time series may be investigated with help of this approach for different objects - sea, river, underground water, etc. In addition to natural periods (equal to or a multiple of a year, a day, a lunar month) the others not so obvious periods may be detected. These periods may be associated with different complicated processes such as anthropogenic factors. SSA method helps to select carefully periods and amplitudes of the oscillations and predict the behavior of the time series.

In this study SSA method is applied for river monthly streamflow investigation and forecasting and for sea level changing near the continents coastline. For purpose of average annual sea level analysis data was taken from Permanent Service for Mean Sea Level (PSMSL) data base [6]. Established in 1933, the PSMSL has been responsible for the collection, publication, analysis and interpretation of sea level data from the global network of tide gauges. It is based in Liverpool at the National Oceanography

adfa, p. 1, 2011.

Centre (NOC), which is a component of the UK National Environmental Research Council (NERC). PSMSL is the global data bank for long term sea level change information from tide gauges and bottom pressure recorders.
In PSMSL a monthly average sea level data is also available. PSMSL data base is available online: http://www.psmsl.org/data/. Change in mean annual level of seas, surrounding the European part of the Eurasian continent is mainly considered in this work, because the Baltic Sea water level time series data is the oldest, longest and most reliable data in the world [6]. The data of long-term observations, received from the most complete and reliable data base, are proceeded and analyzed.
For river monthly streamflow studying data from station on Timpton River, which is 644 km long and situated in Siberia (Russian Federation) was used. Data of 36-year series continuous observation of monthly average river water discharge for given analysis was collected in the line gouge of the designed Nizne-Timpton hydroelectric power station. The purpose is to extracts the trend component, periodical components and noise, as well as the forecasting of trend and each components.
Analysis and forecasting of hydrology indicators - sea water level and river water discharge - are presented as examples of SSA technical application. By the same way various hydrological time series may be investigated, such as water temperature, water flow, underground water level, precipitations, etc.

2 Mathematical methods

The objective of this work is the processing of hydrological time series: smoothing, trends and periodical components extracting, the behavior of each components prediction. Due to SSA method developing many computer programs are created for time series analysis during last years: “Caterpillar”- SSA and RSSA package in St-Petersburg University [3], SSA-MTM Toolkit for Spectral Analysis, kSpectra Toolkit [11], different MATLAB applications.
In this work Mathcad 15.0 programs are applied as easy-to-use environment for engineering and scientific research [8], [9]. Mathcad has a large number of built-in functions, including matrix, and programming tools. The program for time series processing by SSA method in Mathcad media was developed according to SSA algorithm described in Golyandina et al. [5]. Three methods of time series smoothing was used and results were compared: singular spectrum analysis (SSA), simple moving average method and using supsmooth function built in Mathcad program.

Singular spectrum analysis method of time series processing

The algorithm of SSA method consists of transformation of one-dimensional series into multi-dimensional by one-parametric translation procedure, research of received multidimensional trajectory by means of principal components analysis (Singular Value Decomposition) and reconstruction of the series in accordance with the chosen principal components. The result of application of this method is the expansion of the time series into sum of simple components: slow trends, periodic or oscillating components and components of noises. The decomposition received by SSA serve as

basis of correct forecasting. SSA procedure is described in [5], application of SSA method to the hydrological time series processing is given in [7].

Moving average method

A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. The threshold between short-term and long-term depends on the application, and the parameters of the moving average will be set accordingly. Mathematically, a moving average is a type of convolution and so it can be viewed as an example of a low-pass filter used in signal processing. When used with non-time series data, a moving average filters higher frequency components without any specific connection to time, although typically some kind of ordering is implied. Viewed simplistically it can be regarded as smoothing the data. In Mathcad there is built-in function movavg(**v**,n), which the smoothed version of the data in vector **v** created by taking a moving average with a window of width n.

"Supsmooth" function method

Built-in Mathcad function for smoothing X-Y data supsmooth(**vy**, **vx**) returns a vector created by the piecewise use of a symmetric nearest neighbor linear least-squares fitting on each element in **vy**, in which the number of nearest neighbors is adaptively chosen. Arguments in this function **v**x is a vector of real numbers with elements in ascending order, **vy** is a vector of real numbers the same length as **vx**. In our case **vy** is the vector of the number of observations $[1, 2, ...i,\ldots N]^T$, **vx** is the vector of average annual sea level time series $[X_1, X_2,...X_i,\ldots.X_N]^T$.

3 Examples of analysis of sea water level changing in coastal

For sea level analysis data were taken from Permanent Service for Mean Sea Level (PSMSL). PSMSL database forms the revised local reference (or RLR) dataset, RLR datum at each station is defined to be approximately 7000 mm below mean sea level. This conventional choice was made many years ago in order to avoid negative numbers in the resulting RLR monthly and annual mean values.
Three examples of time series is presented: with sea level increasing (Cuxhaven-2 in Germany on the North Sea coast), with constant sea level (Kungholms Fort in Sweden on the Baltic Sea) and with sea level decreasing (Stockholm in Sweden on the Baltic Sea coast). The results of time series processing are presented below in Fig. 1-11. For the graphs in these figures the agreed notations are: X_i - initial series, SSA_tr - its trend calculated in accordance with SSA method, mv_av_i – trend in accordance with moving average method, $Supsm_i$ – in accordance with supsmooth function.

Annual mean sea level in Cuxhaven-2 (Germany) from 1843 till 2016

Cuxhaven-2 tide gauge (station ID:7) is situated on the coast the North Sea, near the port of Hamburg; it is one of the oldest station in the world. Period of the observation is from 1843 till 2016. The completeness of report of Cuxhaven-2 tide gauge is equal 100%, series length N=174. Sea level changing during 174 years is 340 mm. In Fig.1 the initial series and its smoothing is given for SSA trend (window length L=87) with two eigentripples and also with help of moving average method and with

"Supsmooth" function method. Range of smoothing interval for moving average method is equal n=10. As it is seen from the graphs, a good coincidence of two methods of time series processing is observed: singular spectrum analysis and supsmooth function method. Moving average method strongly depends on range of smoothing interval.

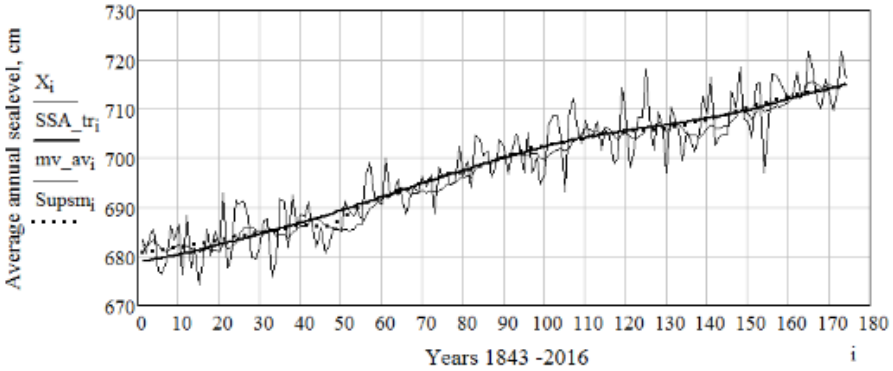


Fig. 1. Cuxhaven-2: initial series and its smoothing (SSA trend with two eigentripples)

Plot in Fig. 2 shows the initial series of annual average sea water level and its trend (with one eigentripple) with trend forecasting for 16 years (shaded area). In Fig. 3 the initial series and its forecasting for 16 years period are presented. Fig. 4 shows the logarithm of eigenvalues in decreasing order; logarithm was introduced because of very large figures. It is clear that the first eigenvalue have the largest value and are able to explain the main part of variance. To determine the size of informative basis the next features of eigenvalue are used: the larger share of eqenvalue in the sum of elements of the eigenvalues vector, the more background information includes a projection on the corresponding eigenvector. It may be supposed that there is noise after second eigenvalue. In Fig. 5 one-dimensional plots of the first, second, third and forth eigenvectors are shown as the example. Analysis of eigenvectors confirm that the first eigenvectors is a trend components (slowly changing component), the last eigenvectors may be regard as noise.

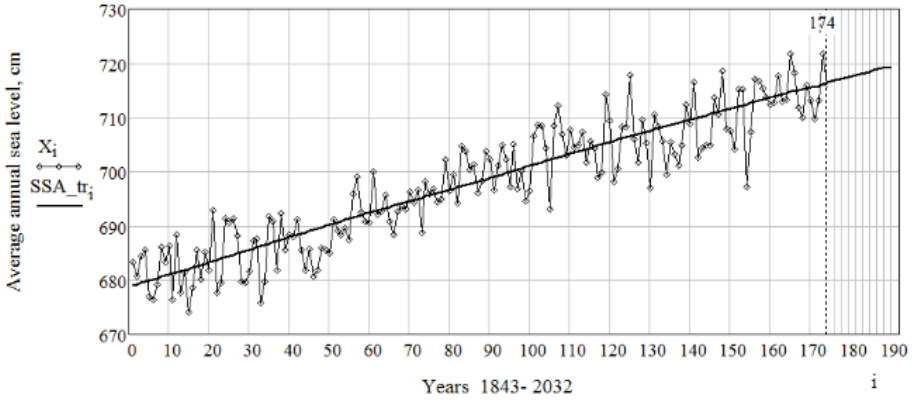


Fig. 2. Cuxhaven-2: initial series and its trend (with one eigentripple) and forecasting to 16 years

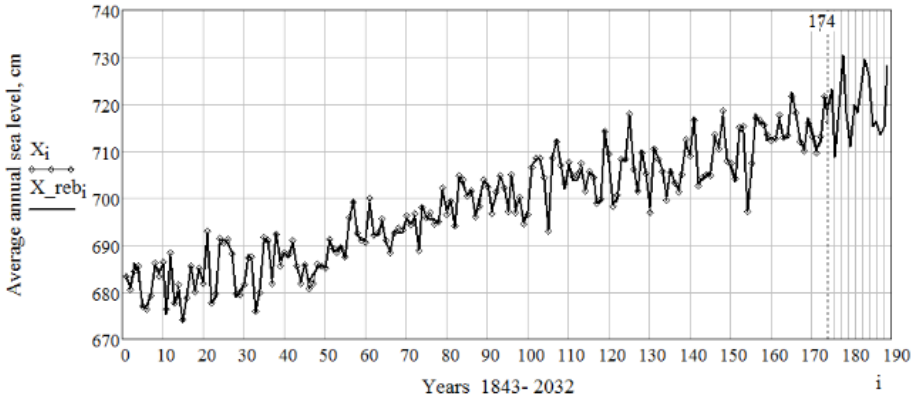


Fig. 3. Cuxhaven-2: initial series and its forecasting for 16 years

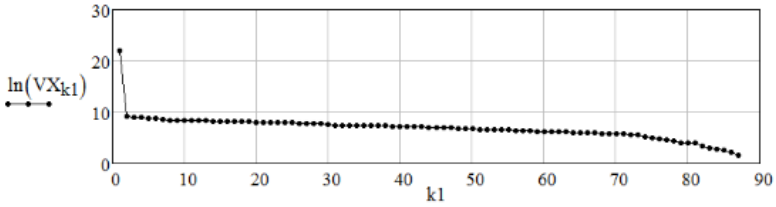


Fig. 4. Singular value spectrum: plot of logarithms of 87 eigenvalues

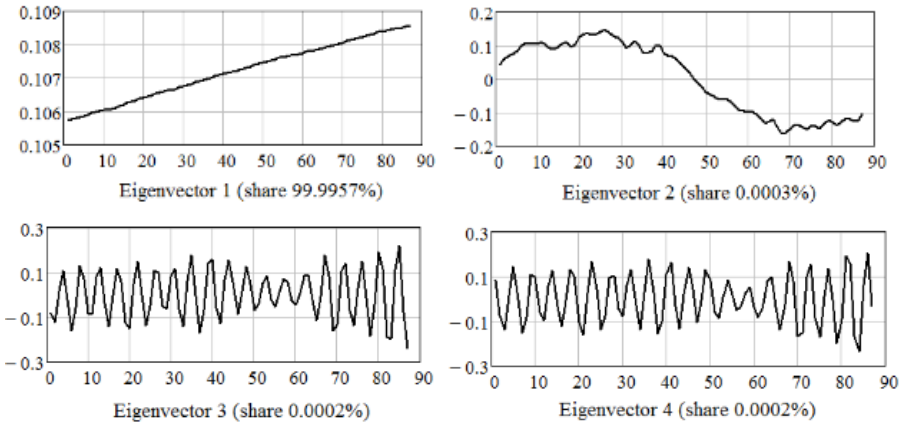


Fig. 5. One-dimensional plots of the first, second, third and fourth eigenvectors

Fig. 6 shows the examples of two-dimensional projection of the eigenvectors.

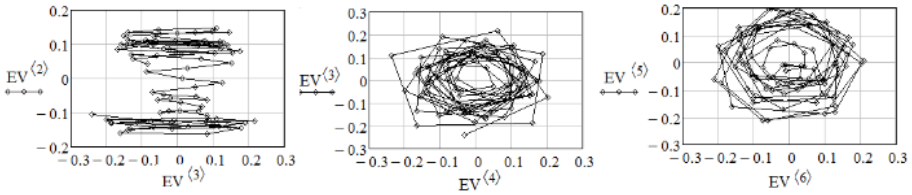


Fig. 6. Two-dimensional plots of eigenvectors

In Fig.6 vectors number is indicated in superscript of function and argument. Analysis of plots in Fig. 6 shows that there are no periodical components.
As it is seen from the plot water level tends to increase. A similar phenomenon of annual average sea level increasing is observed in Tallinn (Estonia), Riga and Liepaja (Latvia), Klaipeda(Lithuania), Swinoujsce (Poland), located on the southern coast of the Baltic Sea.

Annual mean sea level in Kungholms Fort (Sweden) from 1887 till 2017

Kungholms Fort (station ID:70) is situated in the northern coast of the Baltic Sea. Period of observation from 1887 till 2017, .the completeness of report of is equal 100%, series length is N = 131, window length L = 66. Fig. 7 shows the initial series and its smoothing results. Range of smoothing interval for moving average method is equal n =10. In Fig. 8 the initial series and its trend with forecasting to 16 years are presented.

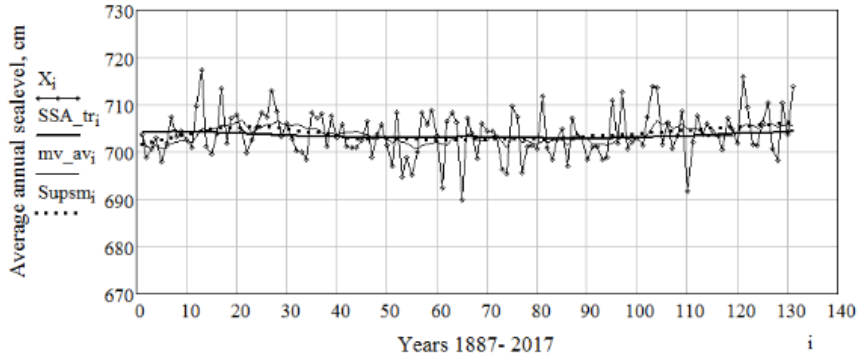


Fig. 7.. Kungholms Fort: initial series and its smoothing (SSA with one eigentripple)

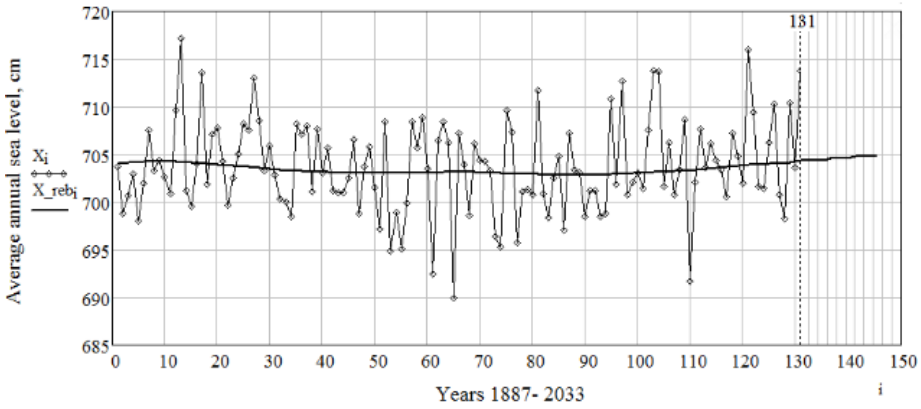


Fig. 8. Kungholms Fort: initial series and its trend (SSA trend with one eigentripple)

Annual mean sea level in Stockholm (Sweden) from 1889 till 2017

Stockholm in (station ID:78) is situated on the northern coast of the Baltic Sea. Series length is N = 129, window length L= 65. Plots in Fig. 9 - 11 show the initial series of annual average sea water level, its smoothing, SSA trend (with one eigentripple) and trend forecasting to 10 years (shaded area). Range of smoothing interval is equal n =10.

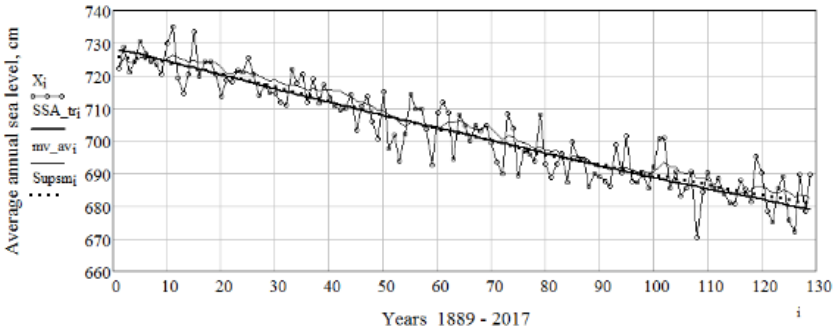


Fig. 9. Stockholm: initial series (line with dot) and its smoothing (SSA trend with one eigentripple)

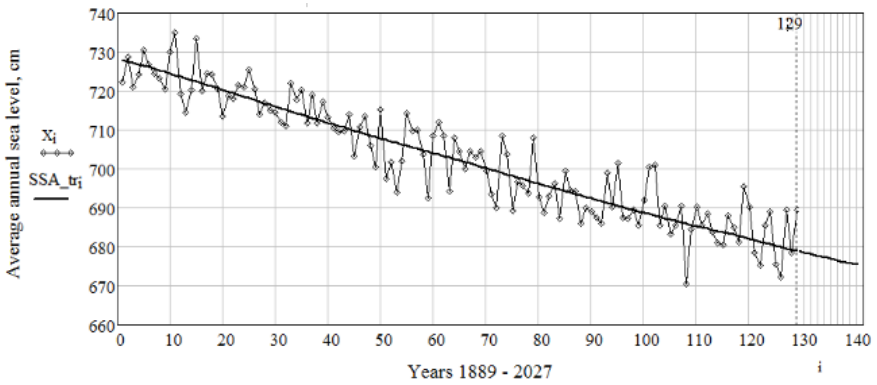


Fig. 10. Stockholm: initial series and its trend (one eigentripple) with forecast on 10 years

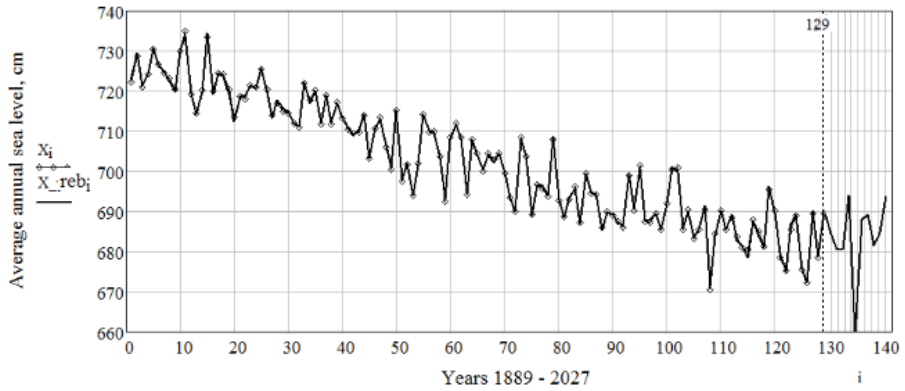


Fig. 11. Stockholm: initial series and its with forecasting to 10 years

The decrease in water level is observed in Ratan (Sweden) and Helsinki (Finland).

Results of sea water levels in costal zones investigation

In table 1 the results of processing of the sea level time series from selected station is given. Data of thirty three stations are presented, the completeness of the most data on

North and Baltic Seas is equal 100%. Baltic Sea water level time series data are the oldest, longest and most reliable data in the world. The completeness of the last data vary between 96-99%. In this work missing data of time series was restored using the MSSA method. Sea level change is calculated as the difference between values of the initial and final points of the trend.

Table 1. Sea level changing

No	Station	Station ID	Sea coast, country	Years of observation		Sea level change, mm	
				period	numb	in period	per year
1	2	3	4	5	6	7	8
Baltic Sea and Gulf of Bothnia							
1	Kungsholms Fort	70	Sweden	1887-2017	131	+35	+0.267
2	Olands Norra Udde	69	Sweden	1887-2017	131	-13	-0.099
3	Stockholm	78	Sweden	1889-2017	129	-457	-3.543
4	Ratan	88	Sweden	1892-2017	126	-955	-7.579
5	Helsinki	14	Finland	1879-2016	138	-342	-2.515
7	Kronstadt		Russia	1835-1993	159	+100	+0.629
8	Klaipeda	118	Lithuania	1898-2011	114	+200	+1.754
9	Swinoujsce	2	Polish	1811-1999	189	+140	+0.741
10	Warnemunde	11	Germany	1856-2016	161	+199	+1.236
11	Wismar 2	8	Germany	1849-2016	168	+197	+1.171
12	Copenhagen	82	Denmark	1889-2012	124	+73	+0.606
North Sea							
13	Cuxhaven-2	7	Germany	1843-2016	174	+340	+1.954
14	Delfzijl	24	Netherland	1865-2016	152	+231	+1.510
15	Ijmuiden	32	Netherland	1872-2016	145	+186	+1.283
16	Maassulis	9	Netherland	1848-2016	169	+293	+1.734
17	Vlissingen	20	Netherland	1862-2016	155	+200	+1.290
18	Oslo	62	Norway	1940-2017	78	-28	-0.359
Norvegian Sea							
19	Kabelvag	45	Norway	1948-2017	70	-93	-0.547
Barenc Sea							
20	Murmansk	684	Russia	1952-2017	63	0	0
North Atlantic Ocean, Bay of Biscay							
21	Brest	1	France	1807-2017	211	+200	+0.962
22	Bilbao	1806	Spain	1993-2016	24	+81	+3.376
Balearic Sea							
23	L'Estarti	1764	Spain	1990-2016	27	+82	+3.037
24	Marseille	61	France	1885-2016	133	+212	+1.571
Adriatic Sea							

1	2	3	4	5	6	7	8
25	Venezia	168	Italy	1923-2000	78	+210	+2.692
26	Trieste	154	Italy	1927-2017	91	+124	+1.363
27	Split	352	Croatia	1955-2014	60	+71	+1.182
28	Dubrovnik	760	Croatia	1956-2014	59	+130	+2.203
Gulf of Mexico							
29	Galevston	161	US	1909-2017	109	+620	+5.636
Nord Pacific Ocean							
30	Seattle	127	US	1899-2017	119	+220	+1.849
Gulf of Farallones							
31	San Francisco	10	United States	1855-2017	163	+280	+1.718
Nord Atlantic Ocean							
32	Boston	235	US	1921-2017	97	+320	+3.229
Tasman Sea							
33	Sydney	65	Australia	1866-1993	108	+50	+0.463

As it is seen from the table 1, there is an overall tendency of the increasing of the mean annual water level up to 1.5-2.0 mm/year. This fact is confirmed by the researchers that reseived data from the another sourses [1], [2], [10].

4 Example of analysis of monthly average water discharge

As the examples of SSA method application the Timpon River discharge were investigated. Timpton River is situated in Saha – Jakutia (Eastern Siberia, Russia), it is 644 km long and watershed area of 44400 km^2. Data of 36-year series continuous observation (from 05.1961 till 05. 1997) was collected in the line gouge of the designed Nizne-Timpton hydroelectric power station. Data of monthly average river water discharge for given analysis was obtained from St.-Petersburg Vedeneeva Scientific Research Institute of hydraulic engineering (Russia). Series length is N = 432, window length L=216. Plots in Fig. 12 shows the initial series of monthly average water discharge, its trend (with one eigentripple) and trend forecasting for 12 month (shaded area).

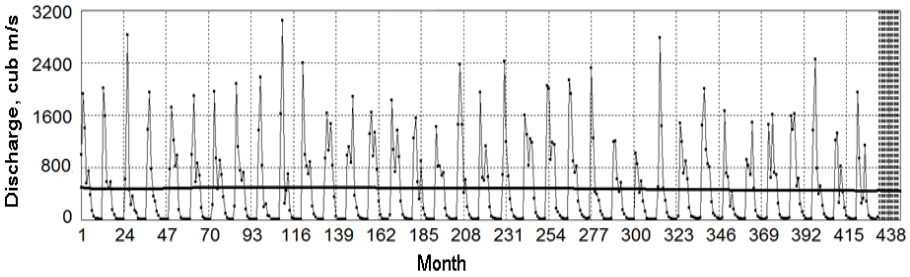


Fig. 12. Monthly average water discharge: initial series and its trend with forecasting for 12 month

Plots in Fig. 13 show the logarithm of eigenvalues in decreasing order; the first nine eigenvalues have the largest value. In Fig. 14 one-dimensional plots of the first, second, third and tenth eigenvectors are shown; Fig. 15 shows the examples of two-dimensional projection of the eigenvectors.

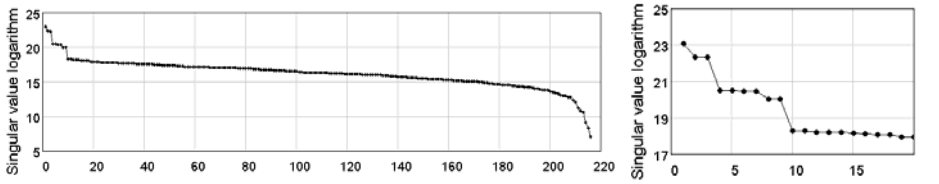


Fig. 13. Singular value spectrum: plot of logarithms of 216 eigenvalues

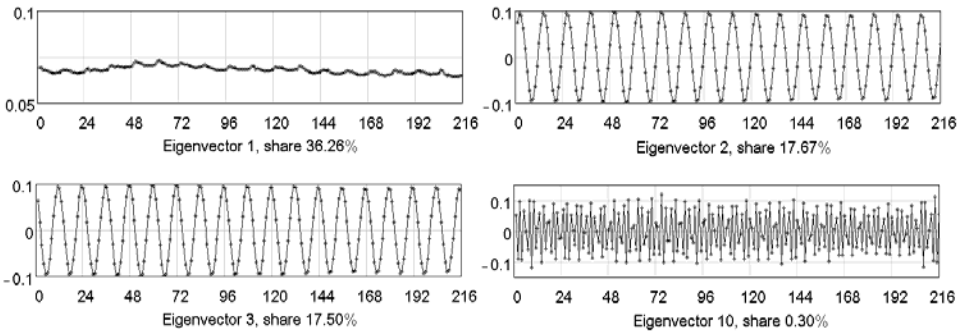


Fig. 14. One-dimensional plots of the 1,2,10 eigenvectors

Analysis of the plots in Fig. 13 – 15 shows that the first eigenvectors is the trend component, the 2-nd and 3-rd eigenvector are periodical with the period 12 month, 4-th and 5-th – periodical with the period 6 month, 6-th and 7-th - with the period 4 month, 8-th and 9-th - with the period 3 month, the last may be refer as noise.

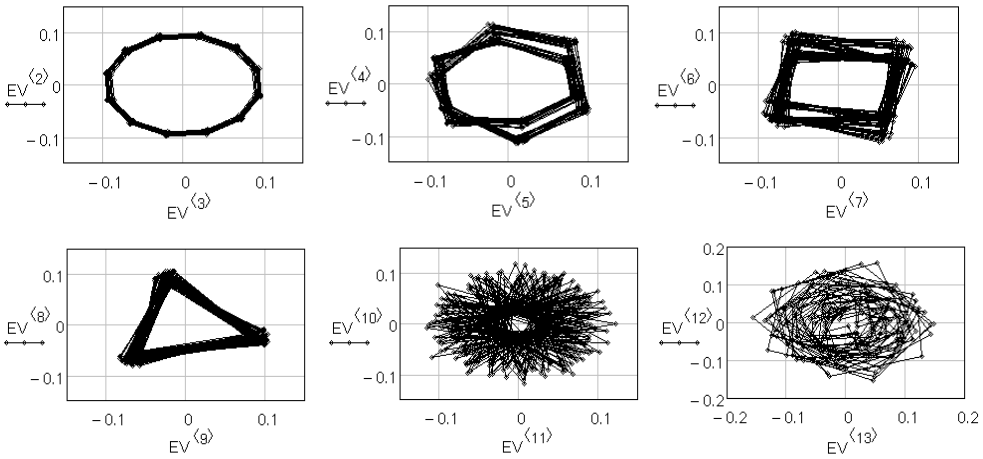
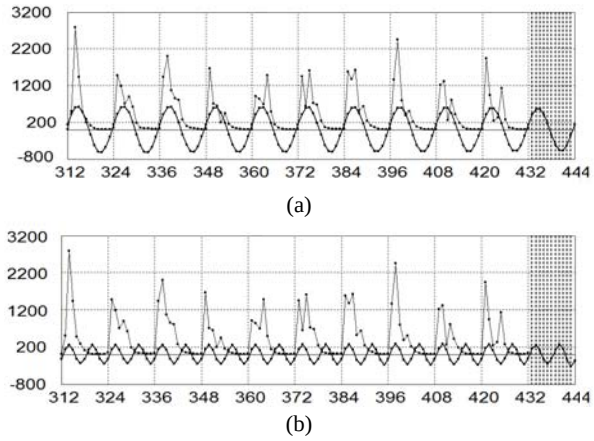


Fig.15. Two-dimensional plots of eigenvectors

Plots of the separated periodical components and their prognosis are presented in Fig.13: with12-month period (b) and with 6-month period (b).



.**Fig.16.** Plots of the separated periodical components and their prognosis: with12-month period (a) and with 6-month period (b)

5 Conclusion

In given work analysis and forecasting of hydrological indicators - sea water level and river water discharge - are presented as examples of SSA technique application. Detailed analysis of Timpton River discharge is fulfilled with graphic illustrations. The important indicators – the contribution flow of each component into the total (in percent) and seasonal components – may be observed from investigation.

The change in mean annual level of the seas surrounding the continents is considered . Tide gauge data obtained from PSMSL database for the most completeness and long-term observations are proceeded and analyzed.

In general, there is an overall tendency of the increasing of the mean annual water level up to 1.5-2.0 mm per year. Change of average annual sea levels vary for various stations. This irregularity is caused by many factors: wind-induced setup, melting of snow and glaciers, production activities of people, groundwater motion in the coast zone, rivers flows.

However, there is a general tendency of the sea level to increase, associated with climate change. The rate of increasing of the water level rises over the past 20 years, which is associated with climate change – general warming. Besides of that, there are trends that cannot be explained by climate change – lowering the water level on the northern coast of the Baltic Sea and the Gulf of Bothnia (Sweden and Finland) up to 7 mm per year. This is probably occurring due to the movement of continents or other geological processes. If current trends will continue over the next 50 years, the raising the water level up to 120-150 mm may occur, which leads to the disfunction of the system of port facilities. Additionally, it is possibly to study the monthly average water level changing over the long term in order to know the seasonal variation of sea level; mean high and mean low sea water level may be estimated by the same way.

By the same way may various hydrological time series can be investigated, such as water temperature, water flow, underground water level, precipitations, etc.

6 References

1. Bogdan, V.I. et al, Mean monthly series of sea level observations (1777-1993) at the Ronstadt gauge. Reports of the Finnish Geodetic Institute, 2000:1, 34pp.
2. Ekman M., The Changing Level of the Baltic Sea during 300 Years: A Clue to Understanding the Earth, Ben Johan's, PQR, Sweden, 2009.
3. Danilov D. & Zhigljavsky A. Principal Components of Time Series: "Caterpillar" Method, St. Petersburg University, Russia, 1997, http://www.gistatgroup.com/gus/
4. Ghil M. et al. Advanced Spectral Methods for Climatic Time Series. Reviews of Geophysics, 40 (1), 2002, pp. **1**-1–**1**-41.
5. Golyandina N., Nekrutkin V. & Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques, CRC Press, USA, 2010.
6. Permanent Service for Mean Sea Level (PSMSL), 2018, "Tide Gauge Data", Retrieved 16 Jul 2018 from http://www.psmsl.org/data/obtaining/
7. Polukoshko S., Hilkevica G., Hydrological Time Series Investigation and Forecasting Using "Caterpillar" – Singular Spectrum Analysis Approach, 14—th GeoConference on Water Resourse. Forest, Marine and Ocean Ecosystems, Conference Proceeding Volume 1, 2014, pp. 389-396.
8. Polukoshko Svetlana, Hilkevica Galina and Gonca Vladimirs, Nonstationary Processes Studying Based on "Caterpillar"- SSA Method. Proceeding of VETOMAC X 2014, University of Manchester, UK, September 9-11, 2014. Mechanism and Machine Science, Volume 23, p. 999-1008,
9. Polukoshko S., Hilkevica G., Gonca V., Non-stationary vibration studying based on singular spectrum analysis. Vibroengineering PROCEDIA, October 2014, Volume 3, pp. 133-138.
10. Simples M.N., Baker T.F., Sea Level Drop in the Mediterranean Sea: An indicator of deep water salinity and temperature changes? Geophysical Research Letters, Vol. 27, pp. 1731-1734, 2000.
11. SSA –MTM Toolkit for Spectral Analysis, http://www.atmos.ucla.edu/tcd/ssa/

Using a naive Bayes classifier to explore the factors driving the harmful dinoflagellate *Alexandrium minutum* dynamics

Wafa Feki-Sahnoun[1]*, Asma Hamza[1], Hasna Njah[2], Nouha Barraj[3], Mabrouka Mahfoudi[1], Ahmed Rebai[4], Malika Bel Hassen[3]

[1] Institut National des Sciences et Technologies de la Mer, Centre de Sfax, Rue Madagascar, BP 1035, Sfax, CP 3018, Tunisie.

[2] Faculté des Sciences Économiques et de Gestion de Sfax, Route de l'Aéroport Km 4 Sfax 3018 Tunisie. Laboratoire de Multimedia, InfoRmationSystems and Advanced ComputingLaboratory, Pôle technologique de Sfax, Route de Tunis Km 10 BP 242 CP 3021, Sfax, Tunisie.

[3] Institut National des Sciences et Technologies de la Mer (INSTM), 28 rue 2 mars 1934, Salammbô 2025, Tunisie.

[4] Centre de Biotechnologie de Sfax, Route Sidi Mansour Km 6, BP 802, Sfax, CP 3019, Tunisie.

Abstract

The blooms of the toxic dinoflagellate *Alexandrium minutum* can be predicted with accuracy derived from knowledge of the main forcing variables. A naive Bayes classifier modeling framework, a member of the Bayesian network family, was used to identify the phytoplankton community, the physical and meteorological parameters involving *Alexandrium minutum* blooms in various sampling sites of the national phytoplankton monitoring program. The proposed model took into account the physical environment effects (salinity, temperature and tide amplitude), meteorological constraints (evaporation, air temperature, insolation, rainfall, atmospheric pressure and humidity), phytoplankton groups (diatoms, dinoflagellates, Cyanobacteria and Euglenophyceae), sampling months and sites on *Alexandrium minutum* blooms. The shift to the highest salinity and water temperature associated with reduced tide and atmospheric pressure were the most favorable conditions regarding the species blooms. The model shows that the species proliferate exclusively in spring and almost in Boughrara lagoon and Djerba Island. It was demonstrated that some diatoms and dinoflagellate appeared during the bloom whereas some others decreased such us Euglenophyceae and Cyanophyceae. This study shows that naive Bayes classifiers can support the management of this ecosystem as they are visually appealing, transparent models that facilitate integration of monitoring data so as to plan for the best disposal options in the treatment of urban and industrial wastes in the Gulf coast.

Keywords: *Alexandrium minutum* blooms, naive Bayes classifier network, physical parameters, meteorological parameters, phytoplankton community, the Gulf of Gabès.

Modeling Global Radiation in Kuwait

S. Al-Awadhi
alawadhidodo@yahoo.com

Department of Statistics and Operation research, Faculty of Science Kuwait University, P.O. Box 5969, Safat 13060, Kuwait

SUMMARY

Two stochastic models that capture the main features of daily exposure of global radiation in Kuwait are proposed. The development of these models is based on removing the annual periodicity and seasonal variation of solar radiation. Thus the daily radiation is decomposed as the sum of trend component and a stochastic component. In many situations, there are dramatic changes in the radiation series through the year due to the condition of the weather as the case of the data from Kuwait. This would affect the accuracy of the model, therefore the series is divided into two regimes: one corresponds to clear days where the value of the global radiation would be normal and the other to non-clear days where the value of global radiation would be very low. Then the trend component is expressed as a Fourier series considering into account such apparent breaks in the series. The stochastic component is first tested for linearity and Gaussianity and found that it does not satisfy these assumptions. Therefore, linear time series model (ARMA modeling) may not be adequate and to overcome this problem, a bilinear time series is used to model the stochastic component of daily global radiation in Kuwait. The method is proposed to consider first fitting a best AR model to the realization and then seeing whether a further reduction in the mean sum of squares can be achieved by introducing extra bilinear terms. The Akaike's Information Criterion (AIC) is used in selecting the best model.

KEY WORDS: Fourier series, ARMA model, bilinear model, stochastic component, MSE, AIC

The predictability of heat-related mortality in Prague, Czech Republic during summer 2015 – A comparison of selected thermal indices

Aleš Urban[1], David M. Hondula[2], Hana Hanzlíková[1], Jan Kyselý[1,3]

[1] Institute of Atmospheric Physics, Czech Academy of Sciences, Prague, Czech Republic
[2] School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA
[3] Faculty of Environmental Sciences, Czech University of Life Sciences, Prague, Czech Republic

email: urban@ufa.cas.cz, kysely@ufa.cas.cz

Abstract. Heat-warning systems are used in many countries to alert the public of elevated health concerns associated with extreme heat, in order to reduce possible impacts on morbidity and mortality. Improved understanding of relationships between meteorological factors and negative health effects can help to refine these measures. The aim of this study is to compare various thermal indices in their ability to predict heat-related mortality in Prague, Czech Republic during the extraordinary summer 2015, based on the temperature-mortality association over 1994–2014. Relatively novel thermal indices – Universal Thermal Climate Index (UTCI) and Excess Heat Factor (EHF) – are compared with more traditional ones (Apparent Temperature (AT), Wet-Bulb Globe Temperature (WBGT), and Physiologically Equivalent Temperature (PET)) to evaluate their suitability for a heat-warning system. After adjusting mortality for long-term trend and seasonality, the relationships between thermal indices and all-cause mortality deviations were estimated by generalized additive models for extended summer seasons (May–September) during 1994–2014, and the resulting models were applied to predict mortality deviations in 2015 based on observed meteorology. The predictions of excess mortality are evaluated in terms of the Critical Success Index.

Although all predictors showed a clear and strong association between thermal conditions and mortality deviations, we found some diversity in their performance. The EHF formula performed best in estimating the intensity of heat waves and magnitude of heat-impacts on excess mortality on the most extreme days, while WBGT14 was most precise in the selection of heat-alert days during the extended summer season, mainly due to a relatively small number of 'false alerts' compared to other predictors. The human-heat-budget based thermal comfort indices (PET and UTCI) were generally less consistent than air temperature and simple indices (WBGT and AT) in predictions of mortality devia-

adfa, p. 1, 2011.

tions as well as in identification of heat-alert days. Air temperature was comparable with or even better than some other (more complex) predictors, possibly also because it requires no additional approximations of input variables enhancing the final model's uncertainty.
Since the main purpose of heat-warning systems is identification of days with an increased risk of heat-related death rather than prediction of exact magnitude of the mortality deviations, WBGT seemed to be a slightly favourable predictor for such a system. Our findings suggest that combination of more techniques within a potential heat-warning system could be useful for a proper understanding of the heat-impact risks on human health.

Keywords: excess mortality; heat stress; thermal indices; warning system; Critical Success Index; Central Europe

Power laws in stock market and fractal complexity of S&P500 and DAX

Anna Krakovská

Institute of Measurement Science, Slovak Academy of Sciences,
Dúbravská Cesta 9, 842 19 Bratislava, Slovak Republic
krakovska@savba.sk

Abstract. In this study the current research evidence about the role of power laws and fractals in stock market data is being reviewed. Then the fractal complexity of stock indexes in US (S&P 500) and Germany (DAX) is estimated. Daily closing prices from 1950 to 2017 are used for calculations.
The results indicate a slightly downward trend in Hurst exponent during the investigated decades, meaning modest decrease in the spectral power-law exponent and increase in the fractal complexity of the indices.
Possible links between the financial crises and changes in complexity are also discussed and questioned.

Keywords: power law, fractal dimension, Hurst exponent, S&P 500, DAX

1 Introduction

Complexity can be generated in a variety of ways. In chaos theory, irregular, hardly predictable behaviour can arise from a few nonlinear differential equations. In agent based models, large number of individual objects governed by a few simple rules are capable of generating remarkably complicated formations. Moreover, complexity may be spatial or temporal - it may involve complicated patterns that do not change over time, or may appear as a surprising time varying behaviour.

In terms of complexity, the stock market is a great place to look at. It is composed of many decision-making individuals with similar motivations. The market is open to the environment, and its dynamics has interesting features. Moreover, the development of this remarkable system is well documented. In the form of stock indexes and price series, we have decades of data to study.

However, if we see the stock market as a system of complex behaviour, then we must call into question the standard random walk model with step sizes that vary according to a normal distribution. However, questioning the normal or log-normal distributions in economics, or in any real-world data, is not easy. Gaussians seem to be all around us. Frequently mentioned examples are distributions of height, weight, measurement errors. We even have the central limit theorem that explains why normal distributions are so common. The theorem

states that the normal distribution arises whenever a large number of independent and identically distributed random variables with finite variances are added together.

In economics, Gaussian distribution of stock returns and the corresponding random walk model are the essence of the efficient market hypothesis [1]. According to this idea, the investors have all information about the expected earnings of the stocks. They should always agree about the right price. The prices can change a little bit after the companies release new estimates of the future earnings. But no such things as bubbles and crises should happen.

Nevertheless, normal distribution may not be as typical as it has long been thought of. Already in 1960 Benoit Mandelbrot pointed out that large movements in prices are much more common than would be predicted from a normal distribution [2]. As an alternative, he suggested the heavy tailed Lévy distributions. The importance of this result was under-valued by the economists until the appearance of modern risk management methods around the 1990s. At that time, the quantitative estimation of the heavy tailed distributions started to be taken much more seriously. Mantegna and Stanley have shown that for the S&P 500 the central part of the distribution corresponds to the Lévy stable process [3]. Scaling behaviour has been observed for time intervals spanning from minutes to weeks. The tails of the distribution deviated from that for a Lévy process and were approximately exponential, ensuring - as one would expect for the price returns - the finite variance of the distribution.

Although heavy-tailed distributions are already accepted in finance, discussions about their origins are still actively going on. For proponents of complex systems, the heavy tails, implying occasional market crashes, and also the formation of bubbles appear as a natural and inherent feature of the system. The first artificial models that confirm this, were designed in the nineties [4]. In the models - computer simulated markets - adaptive learning agents (traders) are trying to predict price movements and buy or sell based on their forecasts. The subsequent change in the price then gets fed back to the traders and is used for further decision about buying or selling. The feedback makes the market a specific dynamic system. Every emergence of obvious predictability and thus the vision of profit opportunity is immediately revealed by a huge number of traders. The massive attempt to use the new discovery immediately leads to canceling or even reversing the pattern. But it is remarkable that in models (and also in reality) sometimes the exact opposite occurs - market participants keep a certain pattern alive. Usually it is a sustained rise in stock prices, known as a bubble. Bubbles are accompanied by an unreasonable increase in the price-earning (P/E) ratio of the shares concerned. The bubbles inevitably crash after some time, but the height of peaks are hard to predict.

Excessive increase in P/E ratio can be caused by unrealistic expectations of future earnings. However, it can also be triggered by herding behaviour. Nate Silver emphasizes in his book [5] that, in the 1960s, only about 15 percent of stocks were held by institutions rather than individuals. By 2007, the percentage had risen to 68 percent. When a trader does not risk his own money, then it may

be quite rational for him to stay with the herd when the market goes up. When the trouble comes, the investor can blame the unpredictability of the market or, at best, admit collective guilt.

Herding behaviour, along with fat tailed distributions, volatility persistence, uncorrelated returns and other stylized facts were replicated by a plethora of agent-based models to date (see e.g. [6] and references therein). What is interesting is that the systems often converge to reality-like complex behavior, even though they have started with very simple initial condition of huge number of agents and a couple of trivial rules. Bak et al. have argued that under rather general conditions the preferred state is the one for which there is no characteristic spatial or temporal scale size [7]. Objects with no spatial scale are fractals (term coined by Mandelbrot in 1970s), and fluctuations with no temporal scale are governed by power laws. For the above phenomenon, in 1980s Bak et al. introduced the term self-organized criticality. In finance it immediately attracted increased attention to the power-law process as to the next candidate for modeling the distribution of returns.

Our variable of interest x has a so called power-law distribution if

$$p(x) \propto 1/x^{\alpha}.$$

Power-law distributions are long-tailed, giving a relatively large probability $p(x)$ to extreme events. They diverge at zero, so there must be a minimum value $x_{min} > 0$ for which the power-law behaviour holds. If $\alpha > 1$, then the median can be computed as $2^{1/(\alpha-1)}x_{min}$ [8]. A power-law process has a well-defined mean only if $\alpha > 2$, and it has a finite variance only if $\alpha > 3$.

Power laws are very common in nature. For example both the gravitational and electrical force decrease inversely with the square of the distance from the mass or charge and thus have no characteristic length scale (until the size of the molecules). Let us also mention the Zipf's law from 1929, valid for the frequency of words in different languages, and for many other phenomena [9]. Another similar law emerged in the 1930s from seismological research of Gutenberg and Richter [10]. Power laws are also ubiquitous in economics. Just remember the Pareto principle from 1896, concerning the income distribution [11].

It is unclear why power laws should be so common. They do not emerge from the generalized central limit theorem as easily as the normal distribution. A generalization of the central limit theorem due to Kolmogorov and Gnedenko states that the sum of a number of random variables with power-law tails tends to a stable distribution only for $1 < \alpha \leq 3$ [12]. If $\alpha > 3$ then the sum converges to the Gaussian distribution. However, most identified power laws in nature are below 3 and they are so common that, sometimes, it would seem that power laws are "the new normal".

Self-organized criticality has been one of the first major attempt to explain the abundance of power laws. Bak et al. declared discovery of a robust mechanism of spontaneous emergence of complexity from simple local interactions - a promising source of ubiquitous complexity. However, the idea has finally shown serious weaknesses and considerable efforts have been made to find new candidates for universal mechanisms for generating power laws [13], [14].

The attraction of power laws probably lies in the fact that processes with power-law distribution are fractals. It means that they are self-similar or self affine and hence scale-free. For example, graphs of share prices over minutes, days, or weeks have a very similar overall form, displaying statistical self-affinity, which means that a rescaled version of a small part of the graph has the same statistical distribution as the larger part. Power-law distributions are accompanying signs of fractals and vice versa. Evidence of inevitable ubiquity of fractals would be impressive.

However, not just the search for a universal mechanism is unsuccessful. We often fail even with a seemingly less ambitious goal of demonstrating that some data follows a power law. The natural procedure is to take the histogram and plot it on a log-log scale. If it looks linear then the estimate of the slope is considered the exponent α. Unfortunately, the data may be well-fitted by a power law, and another function might be an even better fit.

One of the latest studies on this topic comes from Broido and Clauset [15]. The authors compare the fitted power-law distributions to alternative heavy-tailed distributions, like the log-normal or the stretched-exponential. The approximations are compared using a likelihood ratio test. The findings of the authors are clear right from the title of the article - Scale-free networks are rare. The provocative title has prompted an immediate response from Albert-László Barabási, one of the advocates of ubiquitous scale-free networks [16]. He argues that the fitting difficulties are not a reason to dismiss the idea of the scale-free networks. Pure power law only emerges in idealized models. In real networks, we have to admit additional effects. Then, the power law tends to coexist with some corrective function, leading to power laws with exponential cutoffs, stretched exponentials, and so on.

The idea behind the Barabási–Albert networks driven by only growth and the so called preferential attachment is that nodes with higher degree receive more new links than nodes with lower degree [17]. Intuitively this makes sense. Web pages with many incoming links are easier to find, so even more new Web pages will link to them. In networks, this "the rich get richer" principle of preferential attachment leads to scale-free degree distributions. From this point of view the stock market can be seen as a network of traded companies (nodes, vertices). In the network, pairs of vertices are linked by edges, if the level of correlation between the corresponding two price series is above a specified threshold.

Is the preferential attachment, resulting in power-law distributions, the right model for the price returns? Or should we prefer one of many other proposed models and mechanisms? It is not obvious how to decide.

Also, it is still not even clear whether it is appropriate to see the price series and indexes as power-law processes and fractals. On the one hand, price returns certainly have some kind of fat-tailed distributions, but verifying the specific power-law distribution could prove to be be infeasible. On the other hand, we have the remarkable indicators given by Mandelbrot: the apparent statistical self-affinity in data and also the volatility persistence that is strongly supportive

of a long memory property. These are typical features of fractal (power-law) processes.

After this relatively extensive introductory review, the rest of the article is organized as follows. In the next two sections we introduce the data sets and the methods used for estimation of the fractal complexity. The fourth section presents and discusses the results. Finally, the findings are summarized in Conclusion.

2 Methods

Although the idea of a power-law distribution in stock prices is still not justified, use of fractal characteristics such as the Hurst exponent [18], has been quite common for some time.

The Hurst exponent H is usually referred to as a measure of persistence but in fact it is also one of the fractal complexity measures. H can be derived directly from the best-known fractal characteristic, which is the fractal dimension D. Fractal dimension determines the irregularity of the signal. It tells us how smooth or rough the trace of the graph is. D can take values between 1 and 2. The more the graph fills the plane the closer it approaches the value of 2.

Let us also mention the possibility of spectral representation of the investigated scale-free process:

$$S(f) \propto 1/f^{\beta},$$

where $S(f)$ is the power spectral density for the frequency components of the signal. Processes with this power-law relationship are called $1/f$ noises.

When we have self-affine time series, some of the properties as D, decay of autocorrelation or power spectrum, persistence, etc. are elegantly connected. Above all, the next relation between the fractal dimension D, decay of autocorrelation γ, Hurst exponent, and spectral decay β holds:

$$D = \frac{2+\gamma}{2} = 2 - H = \frac{5-\beta}{2} \tag{1}$$

for $1 \leq \beta \leq 3$ [19], [20].

The consequence of (1) is that if you estimate one of the characteristics, you also get the others. For example, in this study, we estimated D and used it to get the estimate of the Hurst exponent and spectral decay.

For computation of D, we chose the Higuchi method, introduced in [21]. The method stood out from our earlier comparison of four different estimation techniques [22].

The Higuchi's method scans fluctuations of the signal by investigating the defined length of the curve for different magnifications of the time axis of the signal. We take time-series $X(1), X(2), ..., X(N)$, and make k lagged time series that start from m-th place ($m = 1, 2, ..., k$) with gap of the size k:

$$X_k^m = X(m), X(m+k), ..., X\left(m + \left[\frac{N-m}{k}\right] k\right)$$

Then Higuchi defines a length of the curve of X_k^m as follows:

$$L_k^m = \sum_{i=1}^{\left[\frac{N-m)}{k}\right]} |X(m+ik) - X(m+(i-1)k| \frac{N-1}{\left[\frac{N-m}{k}\right] k^2}$$

When the length of the curve is calculated (and normalized) for every m and k, we get an $L(k)$ as the mean of all lengths L_k^m. If $L(k) \sim k^{-D}$, then the curve is a fractal with dimension D.

To estimate the Higuchi dimension, we used a Matlab code shared by J. Monge-Álvarez [23].

Real data, however, can only be investigated over a restricted range of scales. We have to determine how high should the value k go. The recommendation is to compute the estimates for increasing k and use the value where the estimates reach a plateau. For the indexes analyzed in this study, the optimal choice was found to be $k = 4$.

After estimating the Higuchi fractal dimension D, we use Eq. (1) to get the corresponding Hurst exponent: $H = 2 - D$.

When we estimate H or other fractal characteristics from the financial series, we actually work with a hypothesis that the given series is a power-law process and thus a fractal.

Finding $H > 0.5$ indicates that the increments of the process are positively correlated, meaning that a high value in the series will probably be followed by another high value (persistence). Variation increases at a faster rate than what is expected for the Gaussian case, so that large jumps are possible, and the series may travel a larger distance than a random walk would imply.

On the other hand, $H < 0.5$ means negative correlation and long-term tendency to switch between high and low values (anti-persistence).

A value of $H = 0.5$ means that the investigated process has independent increments (e.g. ordinary Brownian motion), or the absolute values of the autocorrelations decay exponentially quickly to zero, unlike the above-mentioned case of typically power-law decay. In stock returns data, $H = 0.5$ can be considered to support the efficient market hypothesis.

3 Data

In this study, the daily closing values of S&P 500 and DAX indexes are examined.

The S&P 500 is an American stock market index based on 500 large companies listed on the New York Stock Exchange or the Nasdaq Stock Market. Daily closing values (17110) of S&P 500 from January 3, 1950 to December 29, 2017 were downloaded from https://finance.yahoo.com.

The DAX is a German stock market index consisting of the 30 major companies trading on the Frankfurt Stock Exchange. Daily closing values (14578) of DAX from January 4, 1960 to December 29, 2017 were downloaded from https://stooq.com.

For the computation of the fractal characteristics, the daily closing values of the indexes were used. For the plot of a histogram and the corresponding fit of a normal density function, ordinary returns expressed as a percentage were used.

4 Results and Discussion

Fig. 1 shows the empirical distribution of the returns of S&P 500 for years 1950–2017. The estimated σ is 0.9608. We can see, that the distribution has fatter tails than predicted by the Gaussian. Events that should theoretically be extremely rare, such as price changes by more than 3 percent (above 3σ), are actually quite common. The result suggests that the market does not follow a random walk.

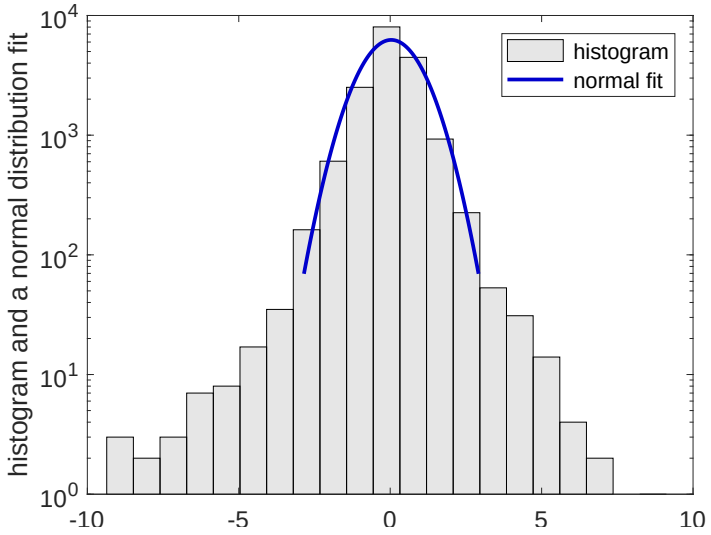


Fig. 1. Histogram for the daily returns of the S&P 500 during the period 1950–2017 compared to the Gaussian normal distribution. In order to amplify the heavy tails of the distributions, semilogarithmic plot is used.

Fig. 2 shows estimates of Hurst exponent for the S&P 500 during the period 1950–2017. Each value of H is estimated from 250 previous daily closing values of the index. Most of the time, the $H > 0.5$ indicates significantly persistent behaviour. However, H shows a slightly downward trend over the years and during the last two decades it oscillates quite close to the value of 0.5 typical for the ordinary random walk. Similar decreasing pattern has been observed when high frequency (1-minute) data were used for S&P 500 index for the period 1983–2009 [24]. The similarity of the results obtained from daily and minute data supports the possible presence of scale-free processes.

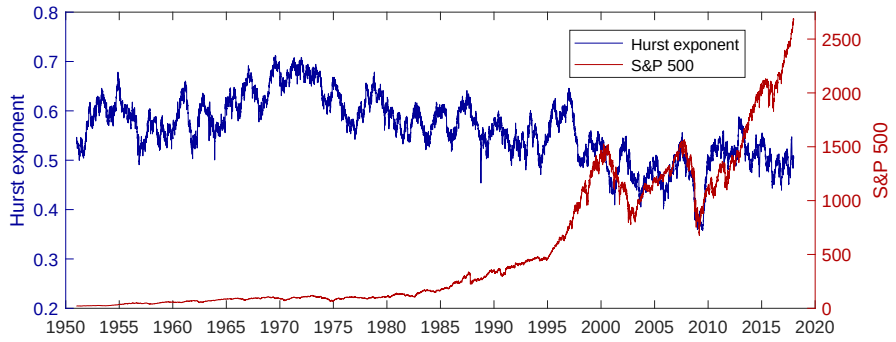


Fig. 2. Hurst exponent for the S&P 500 during the period 1950–2017. Each value of H is estimated from the previous year of daily closing values.

Fig. 3 shows estimates of Hurst exponent for the German DAX during the period 1960–2017. Again, each value of H is estimated from 250 previous daily closing values of the index. Also in this case the Hurst exponent shows a moderate downward trend over the investigated years.

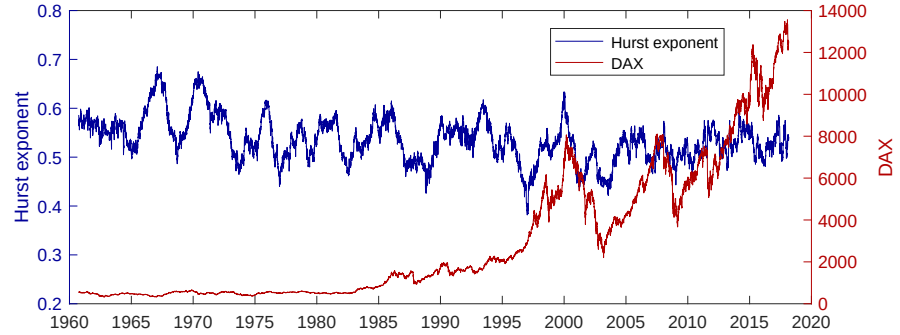


Fig. 3. Hurst exponent for the DAX during the period 1960–2017. Each value of H is estimated from the previous year of daily closing values.

In case of S&P 500, the highest values of H were found for the years 1971–1972. Left part of the Fig. 5 shows a closer look at these years. The index is relatively smooth and persistent. The estimation of the Higuchi dimension is $D \approx 1.3$, so based on Eq. (1), $H \approx 0.7$ and the signal resembles $1/f$ noise with an exponent $\beta \approx 2.4$.

The lowest values of H were found for the turbulent years of 2008–2009. Right part of the Fig. 6 shows S&P 500 and H for these two years. The index here is mostly anti-persistent and it looks more rough and space-filling than in the previous case of 1971–1972. The corresponding estimates of the fractal complexity measures are $D \approx 1.6$, $H \approx 0.4$, $\beta \approx 1.8$.

Can we learn more from the course of the Hurst exponent?

The most important question is, of course, whether H can help to predict the coming crash. By stock market crash we understand panic selling and double-digit percentage losses in a stock market index over a period of few days. Although long-term market predictions are undoubtedly illusory, some authors believe that limited short-term trend predictions are possible [25], [26].

To discuss this topic, let us look closer at the Hurst exponent's course over some specific 2-year periods.

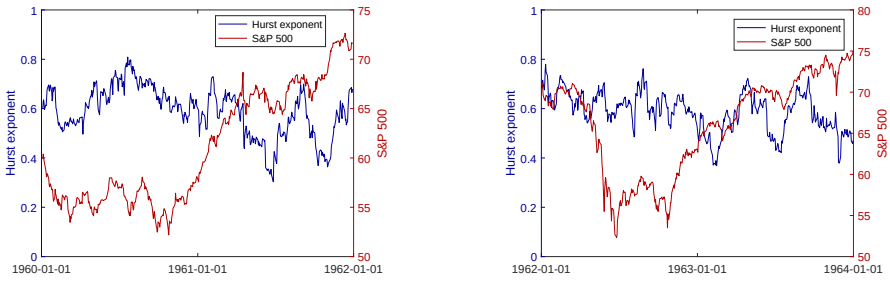


Fig. 4. Hurst exponent (blue) and the S&P 500 (red) during the period 1960–1961 (on the left) and during 1962–1963 (on the right). Each value of H is estimated from 50 previous days.

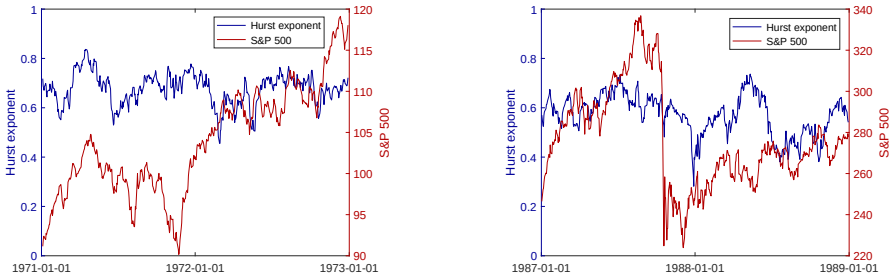


Fig. 5. Hurst exponent (blue) and the S&P 500 (red) during the period 1971–1972 (on the left) and during 1987–1988 (on the right). Each value of H is estimated from 50 previous days.

The right part of Fig. 6 shows estimates of Hurst exponent for the S&P 500 around the financial crisis in September 2008. In the first half of 2008, the stock market was at its peak with $H > 0.5$ suggesting a slight persistence in the data. In such situation, the short-term predictability of the market can be increased, traders may become confident about their strategies and the liquidity, but also the vulnerability of the market grows. What seems to be interesting is

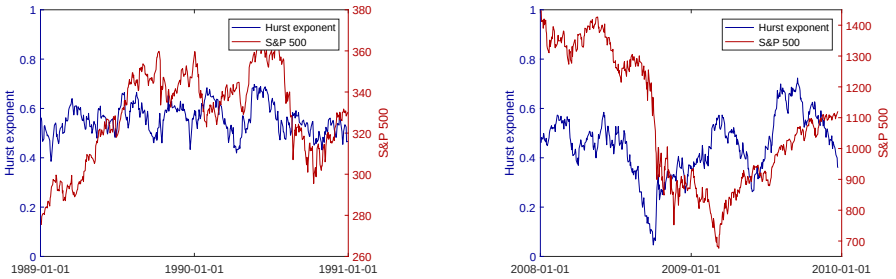


Fig. 6. Hurst exponent (blue) and the S&P 500 (red) during the period 1989–1990 (on the left) and during 2008–2009 (on the right). Each value of H is estimated from 50 previous days.

that well before the crash H began to fall sharply. Unusually low values of the exponent ($H < 0.3$), shortly before the crash, could be interpreted as the onset of strong anti-persistence and therefore extreme nervousness in the market. Do we see a useful prognostic pattern here? Unfortunately, analysis of other similar situations does not confirm this.

Let us see, for example, the strong market decline in 1987. On Black Monday (October 19, 1987), the S&P 500 dropped more than 20%. The crash came after years of strong economic optimism and unprecedented market increase. Estimates of the Hurst exponent for period 1987–1988 can be seen on the right graph of Fig. 5. Similarly, as in 2008, we can see the stock market at its peak with $H > 0.5$ suggesting persistence in data. However, this time, no warning sign of nervousness in the form of a falling Hurst exponent is present.

In other situations, on the contrary, the market looks nervous in the sense that it is anti-persistent, but it still rises. See for example Fig. 4, with periods of rising market, accompanied by H well below 0.5. Another significant example is the rapid market growth in 2017 with S&P 500 booming and anti-persistent at the same time.

5 Conclusion

In this study, we estimated fractal complexity characteristics from daily closing values of US and German stock market indexes. Estimates were mostly found in the following ranges:

$1.3 - 1.6$ for the fractal dimension D,
$0.4 - 0.7$ for the Hurst exponent H,
$1.8 - 2.4$ for the spectral decay β.

In case of S&P 500, the results showed change from the strongly persistent behaviour until 1970s to mostly anti-persistent behaviour of the last two decades.

Then we looked for patterns that are potentially useful for predicting forthcoming danger of stock collapses. However, the prognostic value of the measures of fractal complexity was not confirmed.

We also discussed the role of the power-law processes in stock market modeling. If we accepted power-law processes as inevitable consequences of the evolution of the market as a large complex system, then we would have to accept that:

- The market is very susceptible to any disturbance. Minor shocks can lead to fluctuations of all large sizes, just as it is with earthquakes and avalanches. Catastrophes can occur for no obvious reason, either due to the internal dynamics of the system or triggered by sudden external events like terrorist attacks. But collapses are fundamentally due to the unstable position and they are unavoidable.
- It is not possible to predict the long-term course of the market.
- It is not possible to get rid of the fluctuations through some sort of economic regulations.

Nevertheless, the most robust and efficient state of the economy could be the one with occasional fluctuations of extreme sizes and durations. It is probably not the best imaginable state, but it might be the best achievable state.

The questions that we continue to ask are: Are power laws and fractals typical for the stock market dynamics? If so, is there a universal mechanism that inevitably leads to these power laws?

An overview of current literature shows that we do not have definitive answers. It seems that there are many different ways that fractals and power laws can arise, and that the long sought universal mechanism may not exist. And we still can not rule out the possibility that ubiquitous power laws (fractals) are just mere idealization. As we mentioned above, many published claims for the existence of power laws are almost surely wrong. Maybe what we see are mostly just heavy-tailed distributions. They still indicate a very different phenomenon from Gaussian or exponential, and force us to look at the rare events in a new way. However, there is no doubt that they are not nearly as exciting as power laws and fractals.

Acknowledgments. This work was supported by the Slovak Grant Agency for Science (grants no. 2/0011/16) and by the Slovak Research and Development Agency (grant no. APVV-15-0295). I would also like to thank Hana Krakovská for her help with the computational part of the paper.

References

1. Fama, E.F.: Efficient capital markets: A review of theory and empirical work. The journal of Finance. 25, 2, 383–417 (1970)
2. Mandelbrot, B.: The Pareto-Lévy law and the distribution of income. International Economic Review. 1, 2, 79–106 (1960)
3. Mantegna, R.N., Stanley, H.E.: Scaling behaviour in the dynamics of an economic index. Nature. 376, 6535, 46 (1995)

4. Palmer, R.G., Arthur, W. B., Holland, J.H., LeBaron, B., Tayler, P.: Artificial economic life: a simple model of a stockmarket. Physica D: Nonlinear Phenomena. 75, 1-3, 264–274 (1994)
5. Silver, N.: The signal and the noise: the art and science of prediction. Penguin, UK (2012)
6. Biondo, A. E., Pluchino, A., Rapisarda, A.: Modeling financial markets by self-organized criticality. Physical Review E. 92, 4, 042814 (2015)
7. Bak, P., Tang, C., Wiesenfeld, K.: Self-organized criticality: An explanation of the 1/f noise. Physical Review Letters. 59, 4, 381–384 (1987)
8. Newman, M. E.: Power laws, Pareto distributions and Zipf's law. Contemporary physics. 46, 5, 323-351 (2005)
9. Zipf, G. K.: Relative frequency as a determinant of phonetic change. Harvard studies in classical philology. 40, 1–95 (1929)
10. Gutenberg, B., Richter, C. F.: Magnitude and energy of earthquakes. Science. 83, 2147, 183–185 (1936)
11. Pareto V.: Cours d'Économie Politique Professé a l'Université de Lausanne. Vol. I, (1896)
12. Kolmogorov, A. N., Gnedenko, B. V.: Limit distributions for sums of independent random variables. Addison-Wesley (1968)
13. Gao, J., Cao, Y., Tung, W. W., Hu, J.: Multiscale analysis of complex time series: integration of chaos and random fractal theory, and beyond. John Wiley & Sons. (2007)
14. Gabaix, X.: Power laws in economics: An introduction. Journal of Economic Perspectives. 30, 1, 185–206 (2016)
15. Broido, A. D., Clauset, A.: Scale-free networks are rare. arXiv preprint arXiv:1801.03400 (2018)
16. Barabási, A.-L.: Love is all you need: Clausets fruitless search for scale-free networks. `https://t.co/4UdXeKXefD` (2018)
17. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science. 286, 5439, 509-512 (1999)
18. Hurst, H. E.: Long term storage capacity of reservoirs. ASCE Transactions. 116, 776, 770–808 (1951)
19. Feder, J.: Fractals (physics of solids and liquids). Plennum, New York (1988)
20. Tsonis, A. A.: Chaos: from theory to applications. Springer. Science & Business Media (1992)
21. Higuchi, T.: Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena. 31, 2, 277–283 (1988)
22. Krakovská, H., Krakovská, A.: Fractal Dimension of Self-Affine Signals: Four Methods of Estimation. arXiv preprint arXiv:1611.06190 (2016)
23. Monge-Álvarez, J.: Higuchi Fractal Dimension - MatLab Function. `http://www.mathworks.com/matlabcentral` (2014)
24. Barunik, J., Kristoufek, L.: On Hurst exponent estimation under heavy-tailed distributions. Physica A: Statistical Mechanics and its Applications. 389, 18, 3844–3855 (2010)
25. Sornette, D.: Why stock markets crash: critical events in complex financial systems. Princeton University Press (2017)
26. Grech, D., Pamua, G.: The local Hurst exponent of the financial time series in the vicinity of crashes on the Polish stock exchange market. Physica A: Statistical Mechanics and its Applications. 387, 16–17, 4299–4308 (2008)

Selection of Geographical Factors Using the Random Forest Analysis Method for Developing Site Index of *Pinus densiflora* stands in Republic of Korea

Hee-Jung Park[1†] [0000-0003-0450-0291], Se-Ik Park[1†] [0000-0002-3873-0557], Hyun-Soo Kim[1][0000-0001-8279-3646], Eun-Seong Lee[1][0000-0001-9984-7532], Hyun-Jun Kim[2,3][0000-0002-7373-1643] and Sang-Hyun Lee[1*][0000-0003-1158-9976]

†Contribute equally

[1] Department of Forest Environmental Science, Chonbuk National University, Jeonju 54896, Republic of Korea
[2] BK21 Plus Eco-Leader Education Center, Korea University, Seoul 02841, Republic of Korea
[3] Department of Environmental Science & Ecological Engineering, Korea University, Seoul 02841, Republic of Korea
leesh@jbnu.ac.kr

Abstract. This study was conducted to establish reasonable forest management plan by developing site index curves about *Pinus densiflora* stands that is major species growing in Jeolla-do in Republic of Korea. A total of 613 *Pinus densiflora* plots and 20m×20m sampling plot was installed for each stand. The altitude height, slope, orientation bearing, soil type, the height and diameter at breast height (1.2m from the ground) of a dominant tree, and the age of trees were measured. After developing site index curves using Chapman-Richards, Schumacher and Gompertz models, top 3 geographical factors were added to asymptote and shape parameters. In results, Gompertz model for *Pinus densiflora* stands was chosen for best model of height and site index model. Also, soil type, parent rock and topography were added in to Gompertz model for *Pinus densiflora* stand as the independent variables, using the random forest analysis method. As result of adding geographical factors to asymptote and shape of Gompertz model for *Pinus densiflora* stand, the precision of the model has increased with decreasing MSE. As hybrid site index model including geographical factors indicated influence growth of *Pinus densiflora*, reasonable forest management plan is determined.

Keywords: hybrid site index, site index model, the random forest analysis method, geographical factors, *Pinus densiflora*

1 Introduction

It is difficult to manage the forest in Republic of Korea due to diverse environmental factors including steep and mountainous terrains. Therefore, it is essential to establish a reasonable forest management plan for managing a forest successfully. Moreover, the judgment of site quality is a critical decision factor for establishing a reasonable forest management plan.

Fast-growing stands need thinning mainly at the beginning of management. On the other hand, tending and fertilization are implemented at the beginning of management for slow-growing stands. Therefore, site quality is an essential element to establish the overall forest management plan and select a silvicultural system [14]. Site index is a direct evaluation method of site quality.

Site index is defined as the height of the dominant tree or co-dominant tree at the base age [6,11,15,1,8]. The most important application of site index is to evaluate the production capacity of a stand by examining the possibility of future height growth [2,3,7,5].

A guide curve method and an algebraic difference equation are commonly used to estimate a site index [13]. A guide curve method is a way to estimate a site index by using the height of a dominant tree at one point in a real stand when a permanent sample point and a stem analysis are not available. An algebraic difference equation is a way to estimate a site index from measurements at fixed interval (generally 5-10 years) of a permanent sample point or at least two measurements of a stem analysis [10].

The growth of trees is highly affected by environmental and geographical factors. Therefore, various studies have been conducted to improve the fitness of the model by applying various environmental and geographical factors to the model. Previous studies selected geographical factors suitable for developing site index prediction model by using correlation analysis between a site index and the geographical factors.

The recent development of information power has increased the number of environmental factors and meteorological factors associated with the site index. Therefore, factors affecting the prediction of site index are increasing. It is very important to select the most influential and predictive variable among these numerous independent variables for estimating the site index. Consequently, the random forest analysis method(RFAM), which is one of the ways to determine the priority of independent variables affecting a dependent variable, has been highlighted in many fields. It is one of data mining and machine learning techniques and it reduces the prediction error by maximizing the randomness based on decision tree. Particularly, it has been shown that RFAM has high predictive power for multidimensional data, which have a large number of explanatory variables. RFAM has, recently, been used widely in the forest ecology field because it can avoid the overfitting problem of the single decision analysis using multiple decision trees [4].

This study selected geographical factors suitable for developing a site index predictive equation by applying RFAM, which is one of data mining techniques, to Pinus densiflora stands, which is major species in Jeolla-do region in Republic of Korea. Moreover, the selected geographical factors, which were selected through RFAM, were used to develop the site index model. It was judged that it would be possible to conduct an analysis that could cope with diverse conditions through this geographical factors selection and develop a more accurate site index model. This study was conducted to establish a reasonable forest management plan based on findings.

2 Materials and Methods

2.1 Materials

This study utilized the 2016 survey data of Pinus densiflora stands growing in Jeollado in Republic of Korea. The study selected 613 Pinus densiflora plots and one 20m×20m sampling plot was installed for each stand. Altitude height, slope, orienstation bearing, soil type, the height and diameter at breast height (1.2m from the ground) of a dominant tree, and age of trees were measured (Table 1). Ages of the low data ranged from 10-60 and 61 for the Pinus densiflora stands. Accordingly, this age limits the site index model for those species stands were developed age ranges from 10-60 years and the average age is 39.8 years. Also, height of the low data, ranged from 3.9-21.6m, represents the average height that is 13.2m. The extent of DBH low data is 6.6 to 47.4cm and it appears 25.0cm average value.

Table 1. Status of Species stands.

Species	Number of plots	Age (years)	Height (m)	DBH (cm)
Pinus densiflora	613	$\frac{39.8}{10-60}$	$\frac{13.2}{3.9-21.6}$	$\frac{25.0}{6.6-47.4}$

2.2 Methods

1. Selection of Priority independent variables

RFAM was used to determine the priority of independent variables affecting a site index and RFAM was conducted using R-Studio. An analysis should be carried out by dividing a variable group using a data set. The data set for RFAM in composed of two species and one variable group (11 geographical factors). The independent variables used in the analysis are shown in Table 2.

Table 2. Independent variables used in analysis.

Variable	Factor	Note
X1	altitude height	measurement value
X2	slope	measurement value
X3	orienstation bearing	measurement value
X4	parent rocks	categorical data
X5	topography	categorical data
X6	location of a slope	categorical data
X7	form of an inclination	categorical data
X8	soil depth	categorical data
X9	dry/wet condition	categorical data
X10	soil class	categorical data
X11	soil type	categorical data

It can be hard to interpret the results of RFAM since it cannot derive an intuitive graph, like the decision tree. Therefore, other indices such as variable of importance index and partial dependence plots are provided to understand the importance (influence) of an explanatory variable to a response variable in number or graphs [16]. This study also used these indices.

This study extracted the importance of independent variables affecting the dependent variables (site index) in *Pinus densiflora* stands using RFAM. It was to select geographical factors used as independent variables and develop a hybrid type site index curve with a high degree of fitness.

X1 (altitude height), X2 (slope), and X3 (orienstation bearing), among geographical factors, have a great influence on the extraction of variable importance and they are already known as important variables. They were excluded from the analysis and sued as default variables. The importance of variables from X4 to X11 was analyzed.

2. Dependent variable

This study used Chapman-Richards, Schumacher, Gompertz models based on the assumption that the stem density does not influence the height growth. Moreover, 30-year was used as a base age, which is an important factor for estimating a site index. Forest stands in Republic of Korea were mostly immature so 20-year was used as a base age. However, currently IV age class is the most dominant class in Republic of Korea [9]. Therefore, this study estimated a site index using 30-year, which is the currently used base age in Republic of Korea.

Table 3. General forms of projection models applied to data.

Model name	Model Form*
Chapman-Richards	$H = \alpha[1 - \exp(-\beta \cdot Age)^{\gamma}]$
Schumacher	$H = \alpha \cdot Age^{\beta} \cdot \exp(\frac{\gamma}{Age})$
Gompertz	$H = \alpha \cdot \exp[-\beta \cdot \exp(-\gamma \cdot Age)]$

* H is dominant tree height, exp is exponential function, α, β, γ are coefficients to be estimated.

3. Statistical method of analysis

a. Mean square errors

The Mean square errors (MSE) is the sum of squared errors (MSE) divided by the degree of freedom. It is calculated by considering the estimation offset and the degree of the model.

b. Patterns of residuals

Residuals should be distributed randomly on the X-axis when the developed regression model is valid. Therefore, this study examined if residuals' distribution meat the homoscedasticity assumption.

3 Results and Discussion

1. Selection of priority independent variables

a. Analysis result

The importance of each geographical factors affecting the site index (the mean height of dominant trees at the age of 30 years) was analyzed. The results showed that soil type (26.1801) had the highest influence, followed by parent rock (11.0121) and topography (10.0496) (Fig. 1).

Conclusion of regression analysis to analyze the relatively important geographical factors included in the site index prediction models for *Pinus densiflora* stand living in the central temperate zone. It showed that factors are parent rocks, soil drainage, soil type, B soil depth and dry/wet condition of A soil, which were similar with the results of this study [12].

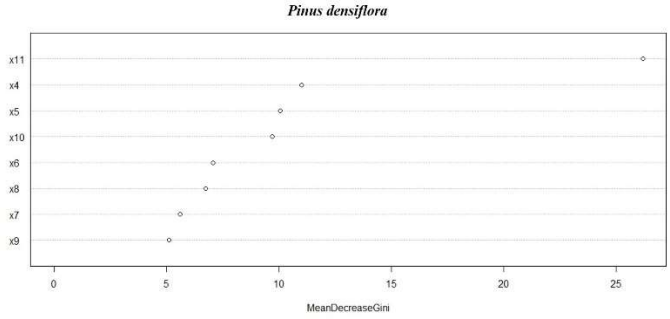


Fig. 1. Variable importance of geographical factors of *Pinus densiflora*.

2. Developing site index

The analysis results showed that the MSE of Chapman-Richards, that of Schumacher and that of Gompertz were 7.1576, 7.1565 and 7.1624, respectively. The MSE of Schumacher model was the smallest. However, it was analyzed that Chapman-Richards and Schumacher models were not statistically reasonable because the 95% confidence limits interval of parameter estimates included "0" range. Therefore, this study developed the site index model by using Gompertz model, which had the MSE value of 7.1624 and the 95% confidence limits interval did not include "0". The developed site index curve agreed with the height of the dominant tree in the corresponding the site index. Moreover, it has the same shape with different magnitudes for the site index ranging between 6 and 18.

Table 4. Non-linear squares summary statistics for 3 models fitted to site index of *Pinus densiflora* stand.

Model name	Source	Degree of freedom	Sum squares	Mean square	F value	Approx Pr>F
Chapman-Richards	Model	3	107823.0	35940.9	5021.35	<.0001
	Error	610	4366.1	7.1576		
	Uncorected total	613	112189.0			
Schumacher	Model	3	107823.0	35941.1	5022.13	<.0001
	Error	610	4365.5	7.1565		
	Uncorected total	613	112189.0			
Gompertz	Model	3	107820.0	35939.9	5017.83	<.0001
	Error	610	4369.1	7.1624		
	Uncorected total	613	112189.0			

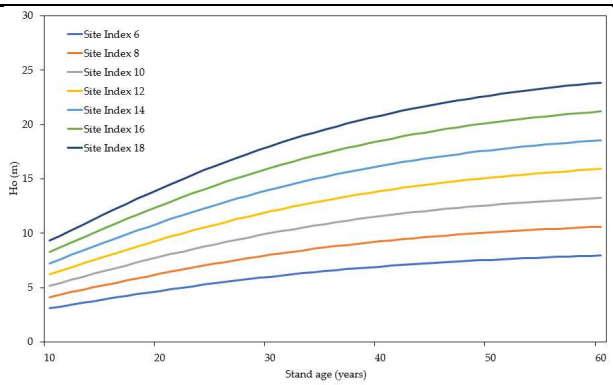


Fig. 2. Site index classification curves for *Pinus densiflora* stand.

3. Developing site index including geographical factors

This study applied the top 3 geographical factors identified from RFAM to the developed site index model for developing a site index model considering geographical factors. The model was divided into the asymptote part and the shape part and the coefficient of each part was controlled. Adding order 1, adding order 1 and 2, adding order 1, 2 and 3 models were divided per asymptote and shape parts and 6 additional models were developed for each species.

a. Adjusting asymptote of Gompertz model

Statistical analyses were conducted by applying geographical factors to the asymptote of the developed Gompertz model. The results showed that the MSE of the basic model was 7.1624, while that with the adding order 1 model,

that with the adding order 1 and 2 model, and that with the adding order 1, 2 and 3 model were 6.1633, 6.0373 and 5.9319. The results indicated that more additional geographical factors decreased MSE value, indicating an improved precision. Moreover, the decrease rate of the error of sum squares compared to the basic equation was 14.0921%, 15.9850% and 17.5871% for the adding order 1 model, that with the adding order 1 and 2 model, and that with the adding order 1, 2 and 3 model, respectively. Additionally, the decrease rate of mean square error was 13.9492%, 15.7084% and 17.1800% for the adding order 1 model, that with the adding order 1 and 2 model, and that with the adding order 1, 2 and 3 model, respectively. The analysis results revealed that it was a good model with high precision because the 95% confidence limits interval of the parameter estimates did not include "0".

Consequently, the results indicated that the asymptote of the site index changed according to the soil type and parent rock. Moreover, tree height growth of a dominant tree varied by the topography (flat land, gentle hill, mountain land). Therefore, the asymptote of the height growth of a dominant tree changed in the model with soil type, parent rock and topography and it was determined that geographical factors affected the precise site index developing.

Table 5. Adjusting asymptote of dominant height model for *Pinus densiflora* stand considering geographical factors.

Model name	Model Forms
Basic	$\mathrm{H} = \alpha \cdot \exp[-\beta \cdot \exp(-\gamma \cdot Age)]$
Adding order 1 (X11)	$\mathrm{H} = (\alpha_1 + \alpha_2 \mathrm{X11}) \cdot \exp[-\beta \cdot \exp(-\gamma \cdot Age)]$
Adding order 1 and 2 (X11, X4)	$\mathrm{H} = (\alpha_1 + \alpha_2 \mathrm{X11} + \alpha_3 \mathrm{X4}) \cdot \exp[-\beta \cdot \exp(-\gamma \cdot Age)]$
Adding order 1, 2 and 3 (X11, X4, X5)	$\mathrm{H} = (\alpha_1 + \alpha_2 \mathrm{X11} + \alpha_3 \mathrm{X4} + \alpha_4 \mathrm{X5}) \cdot \exp[-\beta \cdot \exp(-\gamma \cdot Age)]$

Table 6. Reduction summary statisics for adjusting asymptote of dominant height model for *Pinus densiflora* stand considering geographical factors.

Input variables	ESS	Reduciton ESS (%)	MSE	Reduction MSE (%)
Basic	4369.1	-	7.1624	-
Adding order 1 (X11)	3753.4	▼ 14.0921	6.1633	▼ 13.9492
Adding order 1 and 2 (X11, X4)	3670.7	▼ 15.9850	6.0373	▼ 15.7084
Adding order 1, 2 and 3 (X11, X4, X5)	3600.7	▼ 17.5871	5.9319	▼ 17.1800

b. Adjusting shape of Gompertz model

Statistical analyses were conducted by applying geographical factors to the shape of the developed Gompertz model. The results showed that MSE 7.1624, 6.2497, 6.1376 and 6.0523 for the basic model the adding order 1 model, that with the adding order 1 and 2 model, and that with the adding order 1, 2 and 3 model, respectively, indicating that MSE decreased with more geographical factors. Moreover, the decreased rate of the error of sum squares compared to the basic model was 12.8882%, 14.5888% and 15.9163% for the adding order 1 model, that with the adding order 1 and 2 model, and that with the adding order 1, 2 and 3 model, respectively. Additionally, the decrease rate of mean square error was 12.7429%, 14.3081% and 15.4990% for the adding order 1 model, that with the adding order 1 and 2 model, and that with the adding order 1, 2 and 3 model, respectively. The analysis results revealed that it was a good model with high precision because the 95% confidence limits interval of the parameter estimates did not include "0".

Consequently, the results indicated that the shape of the site index changed according to the soil type, parent rock and topography, which were high priority variables, change the shape of the model and they were geographical factors influencing the growth pattern of a dominant tree. Therefore, it was determined that the site index model reflecting the soil type, parent rock and topography would improve the precision of the site index model.

Table 7. Reduction summary statisics for adjusting shape of dominant height model for *Pinus densiflora* stand considering geographical factors.

Input variables	ESS	Reduciton ESS (%)	MSE	Reduction MSE (%)
Basic	4369.1	-	7.1624	-
Adding order 1 (X11)	3806.0	▼ 12.8882	6.2497	▼ 12.7429
Adding order 1 and 2 (X11, X4)	3731.7	▼ 14.5888	6.1376	▼ 14.3081
Adding order 1, 2 and 3 (X11, X4, X5)	3673.7	▼ 15.9163	6.0523	▼ 15.4990

4 Conclusions

This study prioritized the independent variables affecting the site index using RFAM scientifically and logically. The results of the analysis showed that these variables were in the order of soil type, parent rock and topography for *Pinus densiflora* stands. Conclusion of regression analysis to analyze the relatively important geographical factors included in the site index prediction models for *Pinus densiflora* stand living in the central temperate zone. It showed that factors are parent rocks, soil drainage, soil type, B soil depth and dry/wet

condition of A soil [14], which were similar with the results of this study and it implied that the growth of *Pinus densiflora* in the central temperate zone was highly affected by soil type, parent rock and topography.

Based on these results, this study developed a site index model for *Pinus densiflora* stand growing in Jeolla-do and deduced that Gompertz model provide the most suitable model. Moreover, it was deduced that adding more geographical factors identified by RFAM to the asymptote and shape of the model decreased the MSE value and improved the precision of the model. As a result, it was determined that soil type, parent rock and topography, which had high priority, changed the asymptote and slope of height growth and they were variables effective for estimating the site index more precisely. The result of this study implied that the site index model with additional geographical factors could improve the precision of it for *Pinus densiflora* stands.

This study analyzed the effects of geographical factors on the growth of *Pinus densiflora* growing in Jeolla-do in Republic of Korea. It is determined that reasonable forest management plan can be established using the result of this study. Additionally, it will be necessary to improve the precision of the site index model by prioritizing the meteorological factors affecting the site index as well as geographical factors.

References

1. Avery, T.E., Bukhart, H.E.: Forest measurements. 5th edn. McGraw-Hill, Boston (2002).
2. Bailey, R.L., Clutter, J.L.: Base-age invariant polymorphic site curves. Forest Science 20(2), 248-259 (1974).
3. Biging, G.S.: Improved Estimates of Site Index Curves Using a Varying-Parameter Model. Forest Science 31(1), 248-259 (1985).
4. Breiman, L.: Random Forests. Machine Learning 45(1), 5-32 (2001).
5. Cieszewski, C.J.: Comparing Fixed- and Variable-Base-Age Site Equations Having Single Versus Multiple Asymptotes. Forest Science 48(1), 303-315 (2002).
6. Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., Bailey, R.L.: Timber management: a quantitative approach. John Wiley & Sons, New York(1983).
7. Dyer, M.E., Bailey, R.L.: A test of six methods for estimating true height from stem analysis data. Forest Science 33(1), 3-13(1987).
8. Husch, B., Beers, T.W., Kershaw, J.A.: Forest mensuration. 4th edn. John Wiley & Sons, New Jersey(2003).
9. Korea Forest Service Homepage, http://www.forest.go.kr/newkfsweb/cop/bbs/selectBoardArticle.do?nttId=3098946&bbsId=BBSMSTR_1064&pageIndex=1&pageUnit=10&searchtitle=title&searchcont=&searchkey=&searchwriter=&searchdept=&searchWrd=&ctgryLrcls=&ctgryMdcls=&ctgrySmcls=&ntcStartDt=&ntcEndDt=&orgId=kfs&mn=KFS_02_03_06, last accessed 2018/4/16.
10. Lee, S.H., Seo, B.S., Park, U.J.: Developing Growth Model of Diameter and Height Growth for Quercus Variabilis in Beonsan Peninsula. Korean Forest Economics Society 9(2), 63-68 (2001).
11. Philip, M.S.: Measuring trees and forests. 2nd edn, CAB International, Wallingford (1994).
12. Shin, M.Y., Won H.K., Lee, S.W., Lee, Y.Y.: Site index equations and estimation of productive areas for major pine species by climatic zones using environmental factors. Korean Journal of Agricultural and Forest Meteorology 9(3), 179-187 (2007).
13. Son, Y.M., Lee, K.H., Chung, Y.G.: Articles : Stand Growth Estimation Using Nonlinear Growth Equations. Korean Society Of Forest Science 86(2), 135-145 (1997).
14. Son, Y.M., Lee, K.S., Yoo, B.O., Pyo, J.K.: Estimation of Site Index for Pinus thunbergii in Southern Region of Korea. Korean Journal of Agricultural and Forest Meteorology 47(6), 119-126 (2013).
15. van Laar, A., Akça A.: Forest mensuration. 1st edn, Cuvillier Verlag, Göttingen (1997).
16. Yoo, J.E.: Random forests, an alternative data mining technique to decision tree. Journal of Educational Evaluation 28(2), 427-448 (2015).

The Non-Stationary Unconstrained BINAR(1) Process with Geometric Marginals.

Sunecher Yuvraj, Jowaheer Vandna, Mamode Khan Naushad, Veerasawmy Isven, and Muslun Azmi.

University of Mauritius
Reduit, Mauritius
{ysunecher@umail.utm.ac.mu, vandnaj@uom.ac.mu, n.mamodekhan@uom.ac.mu, isvenveerasawmy@gmail.com , azmi0314@yahoo.com }

Abstract. It is challenging to analyze non-stationary unconstrained bivariate integer-valued autoregressive time series process of order 1 (BINAR(1)) due to the complex cross-correlation structure between the series. This paper considers a novel non-stationary unconstrained BINAR(1) with Geometric marginals where the counting series are commonly influenced by some time-dependent covariates. The Generalized Quasi-Likelihood (GQL) approach is used to estimate the regression and dependence parameters. Monte Carlo simulations are also presented.

Keywords: Non-Stationary; Unconstrained; BINAR(1); GQL; Geometric

1 Introduction

In the recent years, several models for analyzing bivariate integer-valued time series of counts have been developed. Originally, Pedeli and Karlis [9, 10] proposed BINAR(1) processes by extending the two classical INAR(1) processes with Poisson (SBINAR(1)P) and Negative Binomial (NB) (SBINAR(1)NB) innovations respectively under stationary moments. In these models, the cross-relation between the counting series was induced by the correlated pair of bivariate Poisson and bivariate NB respectively, making the bivariate series constrained. In a later research, these authors extended the BINAR(1)s to a full or unconstrained BINAR(1) under same condition of stationarity (FSBINAR(1)P) where the cross-relation was formed by the correlated pair of bivariate Poisson innovations and by the relation between observations from one series with previous-lagged observations of the other series via the binomial thinning mechanism [17, 22]. However, under this correlation set-up, the counting series were proved to be marginally Hermite. Notably, SBINAR(1)NB and FSBINAR(1)P could cater for over-dispersion which is a common phenomenon in real-life time series data.
In addition to these two over-dispersed series models, Ristic et al. [15] and Nastic et al. [8] constructed the stationary BINAR(1) with Geometric marginals (FSBINAR(1)GEOM) where the cross-correlation between the series was borne by the

relation between current series observations with previous-lagged series observations via the NB thinning operator while assuming the innovation series that were mixed Geometrics to be independently distributed. FSBINAR(1)GEOM has a simpler structure than SBINAR(1)NB and FSBINAR(1)P and interestingly, Nastic et al. [8] demonstrated that FSBINAR(1)GEOM yield better AICs than the other corresponding over-dispersed BINAR(1) models. However, FSBINAR(1)GEOM was only built under stationary moment assumptions which defies its purpose in analyzing real-life over-dispersed series that are non-stationary due to time-dependent covariate effects.

As far as the non-stationarity is concerned, Mamode Khan et al. [5] has proposed a BINAR(1) with Poisson innovations (NSBINAR(1)P) where the counting series were under the influence of some common time-varying covariates. In a similar trend, Sunecher et al. [19] constructed the BINAR(1) with NB innovations to model the over-dispersed counting series (NSBINAR(1)NB). Notably, under both model developments, the inter-relation between the series was borne by the correlated pair of bivariate Poisson and bivariate NB innovations respectively. However, there is not yet any paper treating the unconstrained correlation structure specification in the non-stationary BINAR(1)s and in particular in the over-dispersed cases.

Following Nastic et al. [8], we propose in this paper to construct a non-stationary BINAR(1) model with Geometric innovations (FNSBINAR(1)GEOM) where we assume a two-way cross-correlation structure as in Pedeli and Karlis [11] and more importantly under different cross-correlation thinning parameters. In this way, FSBINAR(1)GEOM may be proved as a special case of this new proposed BINAR(1). In fact, in Nastic et al. [7], the authors illustrated that under non-stationary environmental states, it is difficult to specify the joint generating function even though using the decomposability property described by Steutel and Van Harn [18]. Such model construction indeed poses some computational challenges, in particular, in estimating the unknown model parameters.

Firstly, under stationary moment conditions, Pedeli and Karlis [10, 11] developed the Conditional Maximum Likelihood (CML) estimating equations but reported some significant numerical and convergence difficulties (See Pedeli and Karlis [12]), especially in the evaluation of the hessian entries. Note importantly, that in Pedeli and Karlis [10, 11], the authors also compared the efficiency of the CML estimates with the Method of Moments (MoM) and conclude that CML yields far more efficient estimates than MoM. On the other hand, with reference to the FSBINAR(1)GEOM, Nastic et al. [8] showed through some Monte-Carlo experiments that CML is very time-consuming and suggest an alternative Least-Square (LS) technique that omits the specification of the likelihood function. In the same context, Nastic et al. [7] noted that the LS estimates are asymptotically equally efficient as CML while the LS technique yields significant lesser non-convergent simulations than CML.

Under NSBINAR(1)P, Mamode Khan et al. [5] demonstrated that the construction of the conditional likelihood equation is very cumbersome as also stated by Nastic et al. [7]. Thence, for the non-stationary bivariate time series models,

Mamode Khan et al. [5] and Sunecher et al. [19] proposed a generalized quasi-likelihood (GQL) approach that depends only on the correct specification of the mean score and the auto-covariance structure. In Mamode Khan et al. [5], the authors showed that GQL yields asymptotically equally efficient estimates as CML with significantly lesser non-convergent simulations. Thus, based on these research findings, it is worth considering solving the unknown parameters of FNSBINAR(1)GEOM via the GQL and check the efficiency of the estimators.
The organization of the paper is as follows: In the next Section, the new model is constructed. In Section 3, the GQL approach for estimating the unknown parameters of the model is implemented. Section 4 focuses on the simulation part where BINAR(1) data with Geometric marginals are generated and the GQL approach is applied to analyze them. The conclusion is provided in the last Section.

2 Developing the Unconstrained Non-Stationary BINAR(1) with Geometric Marginals

The BINAR(1) model in matrix form is specified as:

$$\boldsymbol{Y}_t = \boldsymbol{A} * \boldsymbol{Y}_{t-1} + \boldsymbol{R}_t \tag{1}$$

with $\boldsymbol{Y_t} = [Y_t^{[1]}, Y_t^{[2]}]^{'}$, $\boldsymbol{R_t} = [R_t^{[1]}, R_t^{[2]}]^{'}$, $\boldsymbol{A} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}$, where $[.]^{'}$ indicates the transpose of the vector.
In the above $Y_t^{[k]}$ is Geometric, $Y_t^{[k]} \sim \text{Geom}(\frac{\mu_t^{[k]}}{1+\mu_t^{[k]}})$, that is $\Pr(Y_t^{[k]} = y) = \frac{\mu_t^{[k]\,y}}{(1+\mu_t^{[k]})^{y+1}}$, $y \geq 0$. The entries of the thinning operation $\boldsymbol{A} * \boldsymbol{Y}_{t-1}$ consists of terms of the form $\rho_{ij} * Y_{t-1}^{[j]}$ where $\rho_{ij} * Y_{t-1}^{[j]} = \sum_{m=1}^{Y_{t-1}^{[j]}} Z_m$ with $Z_m \sim \text{Geom}(\frac{\rho_{ij}}{1+\rho_{ij}})$.
The moments of $R_t^{[k]}$ are yet to be determined. In the above,

$$\text{Corr}(R_t^{[1]}, R_{t'}^{[2]}) = \begin{cases} \kappa_{12,t} & t = t^{'}, \\ 0 & t \neq t^{'}. \end{cases} \tag{2}$$

Referring to Ristic et al. [14] and Silva and Oliveira [16], the following Lemma may be proved easily.
Lemma 1 :

1. $E(\rho_{ij} * Y_{t-1}^{[j]}) = \rho_{ij} E(Y_{t-1}^{[j]})$.
2. $\text{Var}(\rho_{ij} * Y_{t-1}^{[j]}) = \rho_{ij}(1+\rho_{ij})E(Y_{t-1}^{[j]}) + \rho_{ij}^2 \text{Var}(Y_{t-1}^{[j]})$.

In deriving the moments of $R_t^{[k]}$, the following corollaries are proved as well:

Corollary 1 :
The pair $< Y_t^{[k]}, R_{t'}^{[k]} >$ are pairwise independent for $t \neq t'$ such that

$$\mathrm{Cov}(Y_t^{[k]}, R_{t'}^{[k]}) = \begin{cases} \mathrm{Var}(R_t^{[k]}) & t = t^{'}, \\ 0 & t \neq t^{'}. \end{cases} \tag{3}$$

Corollary 2 :
$< Y_t^{[k]}, R_{t'}^{[k']} >$ are also mutually independent for $t \neq t'$ and $k \neq k'$ with

$$\mathrm{Cov}(Y_t^{[k]}, R_{t'}^{[k']}) = \begin{cases} \mathrm{Cov}(R_t^{[k]}, R_t^{[k']}) & t = t^{'}, \\ 0 & t \neq t^{'}, k, k' = 1, 2. \end{cases} \tag{4}$$

Hence, $E(R_t^{[1]}) = \lambda_t^{[1]} = \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}$ and $E(R_t^{[2]}) = \lambda_t^{[2]} = \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]}$ and using Lemma 1,

$$\begin{aligned} \mathrm{Var}(R_t^{[1]}) = {} & \mu_t^{[1]}(1+\mu_t^{[1]}) - \rho_{11}(1+\rho_{11})\mu_{t-1}^{[1]} - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}(1+\rho_{12})\mu_{t-1}^{[2]} \\ & - \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]}) - 2\rho_{11}\rho_{12}\mathrm{Cov}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) \end{aligned} \tag{5}$$

and

$$\begin{aligned} \mathrm{Var}(R_t^{[2]}) = {} & \mu_t^{[2]}(1+\mu_t^{[2]}) - \rho_{21}(1+\rho_{21})\mu_{t-1}^{[1]} - \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{22}(1+\rho_{22})\mu_{t-1}^{[2]} \\ & - \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]}) - 2\rho_{21}\rho_{22}\mathrm{Cov}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]}). \end{aligned} \tag{6}$$

which may be re-expressed as:

$$\begin{aligned} \mathrm{Var}(R_t^{[1]}) = {} & E(R_t^{[1]}) + \{{\mu_t^{[1]}}^2 - \rho_{11}^2\mu_{t-1}^{[1]} - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}^2\mu_{t-1}^{[2]} \\ & - \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]}) - 2\rho_{11}\rho_{12}\mathrm{Cov}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]})\} \end{aligned} \tag{7}$$

and

$$\begin{aligned} \mathrm{Var}(R_t^{[2]}) = {} & E(R_t^{[2]}) + \{{\mu_t^{[2]}}^2 - \rho_{21}^2\mu_{t-1}^{[1]} - \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{22}^2\mu_{t-1}^{[2]} \\ & - \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]}) - 2\rho_{21}\rho_{22}\mathrm{Cov}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]})\}. \end{aligned} \tag{8}$$

These moments are sufficient to conclude that the marginal distribution of $R_t^{[k]}$ has a rather complex probability function structure.

Let us define $\boldsymbol{\Sigma}_{h,t} = \begin{bmatrix} \mathrm{Cov}(Y_t^{[1]}, Y_{t+h}^{[1]}) & \mathrm{Cov}(Y_{t+h}^{[1]}, Y_t^{[2]}) \\ \mathrm{Cov}(Y_t^{[1]}, Y_{t+h}^{[2]}) & \mathrm{Cov}(Y_t^{[2]}, Y_{t+h}^{[2]}) \end{bmatrix}$. From Pedeli and Karlis [11] and Ristic et al. [15], it was shown that $\boldsymbol{\Sigma}_{h,t} = \boldsymbol{A}^h \boldsymbol{\Sigma}_{0,t}$ and hence

$$\begin{aligned} \mathrm{Cov}(Y_t^{[1]}, Y_t^{[2]}) &= \mathrm{Cov}(\rho_{11} * Y_{t-1}^{[1]} + \rho_{12} * Y_{t-1}^{[2]} + R_t^{[1]}, \rho_{21} * Y_{t-1}^{[1]} + \rho_{22} * Y_{t-1}^{[2]} + R_t^{[2]}) \\ &= (\rho_{11}\rho_{22} + \rho_{12}\rho_{21})\mathrm{Cov}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) + \rho_{11}\rho_{21}\mathrm{Var}(Y_{t-1}^{[1]}) + \rho_{22}\rho_{12}\mathrm{Var}(Y_{t-1}^{[2]}) + \mathrm{Cov}(R_t^{[1]}, R_t^{[2]}) \\ &= (\rho_{11}\rho_{22} + \rho_{12}\rho_{21})\mathrm{Cov}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) + \rho_{11}\rho_{21}(\mu_{t-1}^{[1]} + {\mu_{t-1}^{[1]}}^2) + \rho_{22}\rho_{12}(\mu_{t-1}^{[2]} + {\mu_{t-1}^{[2]}}^2) \\ &\quad + [\kappa_{12,t}\sqrt{\mathrm{Var}(R_t^{[1]})}\sqrt{\mathrm{Var}(R_t^{[2]})}] \end{aligned} \tag{9}$$

Note that under stationary assumptions and with $\alpha = \rho_{11} = \rho_{12}$, $\beta = \rho_{21} = \rho_{22}$, $\mu = \mu_{t-1}^{[1]} = \mu_{t-1}^{[2]}$ and $\kappa_{12,t} = 0$, equation (9) simply reduces to the covariance formula in Ristic et al. [15], that is $\text{Cov}(Y_t^{[1]}, Y_t^{[2]}) = \frac{2\alpha\beta}{1-2\alpha\beta}\mu(1+\mu)$.
Under $\rho_{12} = \rho_{21} = 0$, the BINAR(1) model (1) reduces to the constrained non-stationary BINAR(1) (NSBINAR(1)GEOM). The moments and cross-covariance expressions (equations (5)-(9)) follow from these substitution.

3 Estimation Method:Generalized Quasi-Likelihood Technique

We consider here the GQL estimating function to estimate the regression and dependence parameters. This approach requires only the correct specification of the score, its corresponding mean function and the exact auto-covariance structure. In general, the GQL function is expressed as

$$\boldsymbol{D_\phi}^{'} \boldsymbol{\Sigma_\phi}^{-1}(\boldsymbol{y_\phi} - \boldsymbol{\mu_\phi}) = 0 \tag{10}$$

where $\text{E}(\boldsymbol{y_\phi}) = \boldsymbol{\mu_\phi}$ and $\boldsymbol{D_\phi}$ is the derivative matrix consisting of the entries $\{\frac{\partial \boldsymbol{\mu_\phi}}{\partial \boldsymbol{\phi}}\}$, $\boldsymbol{\Sigma_\phi}$ is the exact or robust auto-covariance representation. Note here that some special cases such as in estimating the mean and over-dispersion parameters where the underlying model is NB [2] or COM-Poisson [6], the implementation of a single GQL may not be sufficient and hence requires more than one GQL equation. In Jowaheer and Sutradhar [3], the score function was altered in such a way that the auto-covariance function $\boldsymbol{\Sigma_\phi}$ constitute of covariance of higher-order entries that necessitates the usage of the multivariate normality working structure of Prentice and Zhao [13]. In general, Sutradhar [20] and Sutradhar et al. [21] established the asymptotic properties of the GQL equation in (10) and showed that the GQL estimators are asymptotically normal and consistent irrespective of the working or exact covariance structure specification.
We firstly describe the GQL to estimate the regression parameters as:

$$\boldsymbol{D_\beta}^{'} \boldsymbol{\Sigma_\beta}^{-1}(\boldsymbol{f} - \boldsymbol{\mu}) = 0 \tag{11}$$

with score vector $\boldsymbol{f} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_t, \ldots, \boldsymbol{f}_{t+h}, \ldots, \boldsymbol{f}_T]$ with$\boldsymbol{f}_t = [Y_t^{[1]}, Y_t^{[2]}]'$ and $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_t, \ldots, \boldsymbol{\mu}_T]$ with $\boldsymbol{\mu}_t = [\mu_t^{[1]}, \mu_t^{[2]}]'$ for $t = 1, 2, \ldots, T$. $\boldsymbol{\Sigma_\beta}$

represents the $(2T \times 2T)$ auto-covariance matrix of the form

$$\begin{pmatrix}
\boldsymbol{\Sigma}_{0,1} & \boldsymbol{\Sigma}'_{1,1} & \boldsymbol{\Sigma}'_{2,1} & \dots & \boldsymbol{\Sigma}'_{t-1,1} & \dots & \boldsymbol{\Sigma}'_{t+h-1,1} & \dots & \boldsymbol{\Sigma}'_{T-1,1} \\
\boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{0,2} & \boldsymbol{\Sigma}'_{1,2} & \dots & \boldsymbol{\Sigma}'_{t-2,2} & \dots & \boldsymbol{\Sigma}'_{t+h-2,2} & \dots & \boldsymbol{\Sigma}'_{T-2,2} \\
\boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{1,2} & \boldsymbol{\Sigma}_{0,3} & \dots & \boldsymbol{\Sigma}'_{t-3,3} & \dots & \boldsymbol{\Sigma}'_{t+h-3,3} & \dots & \boldsymbol{\Sigma}'_{T-3,3} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\boldsymbol{\Sigma}_{t-1,1} & \boldsymbol{\Sigma}_{t-2,2} & \boldsymbol{\Sigma}_{t-3,3} & \dots & \boldsymbol{\Sigma}_{0,t} & \dots & \boldsymbol{\Sigma}'_{t+h-t,t} & \dots & \boldsymbol{\Sigma}'_{T-t,t} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\boldsymbol{\Sigma}_{t+h-1,1} & \boldsymbol{\Sigma}_{t+h-2,2} & \boldsymbol{\Sigma}_{t+h-3,3} & \dots & \boldsymbol{\Sigma}_{t+h-t,t} & \dots & \boldsymbol{\Sigma}_{0,t+h} & \dots & \boldsymbol{\Sigma}'_{T-t,t} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\boldsymbol{\Sigma}_{T-1,1} & \boldsymbol{\Sigma}_{T-2,2} & \boldsymbol{\Sigma}_{T-3,3} & \dots & \boldsymbol{\Sigma}_{T-t,t} & \dots & \boldsymbol{\Sigma}_{T-t+h,t} & \dots & \boldsymbol{\Sigma}_{0,T}
\end{pmatrix}_{(2T\times 2T)} \tag{12}$$

where $\boldsymbol{\Sigma_{h,t}}$ is specified as in Section 2.

The derivative matrix $\boldsymbol{D_\beta} = [\boldsymbol{D}_1, \boldsymbol{D}_2, \dots, \boldsymbol{D}_t, \dots, \boldsymbol{D}_T]'$ with $\boldsymbol{D}_t = \begin{pmatrix} \frac{\partial \mu_t^{[1]}}{\partial \boldsymbol{\beta}^{[1]}} & 0 \\ 0 & \frac{\partial \mu_t^{[2]}}{\partial \boldsymbol{\beta}^{[2]}} \end{pmatrix}_{2p\times 2}$ and $\frac{\partial \mu_t^{[k]}}{\partial \boldsymbol{\beta}^{[k]}} = [\frac{\partial \mu_t^{[k]}}{\partial \beta_j^{[k]}}]_{j=1:p}$ where $\frac{\partial \mu_t^{[k]}}{\partial \beta_j^{[k]}} = \mu_t^{[k]} x'_{tj}$.

As for the dependence parameter $\kappa_{12,t}$, they are estimated using the method of moments as follows:

$$\kappa_{12,t} = \frac{\tilde{\text{Cov}}(Y_t^{[1]}, Y_t^{[2]}) - (\hat{\rho}_{11}\hat{\rho}_{22} + \hat{\rho}_{12}\hat{\rho}_{21})\tilde{\text{Cov}}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) - \hat{\rho}_{11}\hat{\rho}_{21}(\hat{\mu}_{t-1}^{[1]} + \hat{\mu_{t-1}^{[1]}}^2) - \hat{\rho}_{22}\hat{\rho}_{12}(\hat{\mu}_{t-1}^{[2]} + \hat{\mu_{t-1}^{[2]}}^2)}{\sqrt{\text{Var}(R_t^{[1]})}\sqrt{\text{Var}(R_t^{[2]})}} \tag{13}$$

where $\hat{\mu}_0^{[k]} = \hat{\mu}_1^{[k]}$, $\tilde{\text{Cov}}(Y_t^{[1]}, Y_t^{[2]}) = \frac{1}{T}\sum_{t=1}^{T}(y_t^{[1]} - \hat{\mu}_t^{[1]})(y_t^{[2]} - \hat{\mu}_t^{[2]})$ and $\tilde{\text{Cov}}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) = \frac{1}{T-1}\sum_{t=2}^{T}(y_{t-1}^{[1]} - \hat{\mu}_{t-1}^{[1]})(y_{t-1}^{[2]} - \hat{\mu}_{t-1}^{[2]})$.

Assuming $\boldsymbol{\beta} = [\boldsymbol{\beta}^{[1]}, \boldsymbol{\beta}^{[2]}]'$ with each $\boldsymbol{\beta}^{[k]}$ a vector of $(p\times 1)$ covariates $[\beta_1^{[k]}, \dots, \beta_j^{[k]}, \dots, \beta_p^{[k]}]'$, the Newton-Raphson (NR) is employed to estimate

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{r+1}^{[1]} \\ \hat{\boldsymbol{\beta}}_{r+1}^{[2]} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_{r}^{[1]} \\ \hat{\boldsymbol{\beta}}_{r}^{[2]} \end{pmatrix} + [\boldsymbol{D_\beta}' \boldsymbol{\Sigma_\beta}^{-1} \boldsymbol{D_\beta}]_r^{-1} [\boldsymbol{D_\beta}' \boldsymbol{\Sigma_\beta}^{-1} (\boldsymbol{f} - \boldsymbol{\mu})]_r \tag{14}$$

Thus, for a given $[\hat{\rho}_{11}, \hat{\rho}_{12}, \hat{\rho}_{21}, \hat{\rho}_{22}, \hat{\kappa}_{12,t}]$ and initial $\boldsymbol{\beta}$, the iterative equation (14) is solved until convergence. $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'$ is shown to follow asymptotically

the normal distribution with mean 0 and covariance matrix matrix specified by $[\boldsymbol{D_\beta}^{'}\boldsymbol{\Sigma_\beta}^{-1}\boldsymbol{D_\beta}]^{-1}$. The derivation of the above asymptotic formula can be seen in Sutradhar et al. [21], Mamode Khan et al. [5] and Sunecher et al. [19].
Using the updated vector for each series, another GQL for the dependence parameter $\boldsymbol{\psi} = [\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}]$ is constructed as:

$$\boldsymbol{D_\psi}^{'}\boldsymbol{\Sigma_\psi}^{-1}(\boldsymbol{Y_\psi} - \boldsymbol{\mu_\psi}) = 0, \tag{15}$$

with $\boldsymbol{Y_\psi} = [Y_1^{[1]}Y_1^{[2]}|Y_0^{[1]}, Y_0^{[2]}, Y_2^{[1]}Y_2^{[2]}|Y_1^{[1]}, Y_1^{[2]}, \ldots, Y_t^{[1]}Y_t^{[2]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}, \ldots, Y_T^{[1]}Y_T^{[2]}|Y_{T-1}^{[1]}, Y_{T-1}^{[2]}]'_{T\times 1}$
with $\boldsymbol{\mu_\psi} = [E(Y_1^{[1]}Y_1^{[2]}|Y_0^{[1]}, Y_0^{[2]}), E(Y_2^{[1]}Y_2^{[2]}|Y_1^{[1]}, Y_1^{[2]}), \ldots, E(Y_t^{[1]}Y_t^{[2]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}), \ldots$
$, E(Y_T^{[1]}Y_T^{[2]}|Y_{T-1}^{[1]}, Y_{T-1}^{[2]})]'_{T\times 1}$. The $T\times T$ covariance matrix $\boldsymbol{\Sigma_\psi}$ constitute of $\mathrm{Var}(Y_t^{[1]}Y_t^{[2]} \mid Y_{t-1}^{[1]}, Y_{t-1}^{[2]})$ along the diagonal and the off-diagonal entries are given by $\mathrm{Cov}(Y_t^{[1]}Y_t^{[2]}Y_{t+h}^{[1]}Y_{t+h}^{[2]} \mid Y_{t-1}^{[1]}, Y_{t-1}^{[2]}, Y_{t+h-1}^{[1]}, Y_{t+h-1}^{[2]})$. Since the entries are of high-ordered moments, the 'working' multivariate normality assumption structure [1, 3, 4, 13] is used to approximate these expressions. Specifically, the entries $\mathrm{Var}(Y_t^{[1]}Y_t^{[2]} \mid Y_{t-1}^{[1]}, Y_{t-1}^{[2]})$ and $\mathrm{Cov}(Y_t^{[1]}Y_t^{[2]}Y_{t+h}^{[1]}Y_{t+h}^{[2]} \mid Y_{t-1}^{[1]}, Y_{t-1}^{[2]}, Y_{t+h-1}^{[1]}, Y_{t+h-1}^{[2]})$ are derived in similar steps as outlined in Sunecher et al. [19] and Jowaheer et al. [2], using the conditional expressions in the Appendix section of this paper.
$\mathrm{Var}(Y_t^{[1]}Y_t^{[2]} \mid Y_{t-1}^{[1]}, Y_{t-1}^{[2]})$

$$\begin{aligned}
&= [\rho_{11}(1+\rho_{11})[Y_{t-1}^{[1]} - \mu_{t-1}^{[1]}] + \rho_{12}(1+\rho_{12})[Y_{t-1}^{[2]} - \mu_{t-1}^{[2]}] + \mu_t^{[1]}(1+\mu_t^{[1]}) - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) \\
&- \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})] \times [\rho_{21}(1+\rho_{21})[Y_{t-1}^{[1]} - \mu_{t-1}^{[1]}] + \rho_{22}(1+\rho_{22})[Y_{t-1}^{[2]} - \mu_{t-1}^{[2]}] + \mu_t^{[2]}(1+\mu_t^{[2]}) \\
&- \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})] + 2\phi_{12,t}^2 + {\mu_t^{[2]}}^2[\rho_{11}(1+\rho_{11})[Y_{t-1}^{[1]} - \mu_{t-1}^{[1]}] \\
&+ \rho_{12}(1+\rho_{12})[Y_{t-1}^{[2]} - \mu_{t-1}^{[2]}] + \mu_t^{[1]}(1+\mu_t^{[1]} - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})] \\
&+ {\mu_t^{[2]}}^2[\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}]^2 + 4{\mu_t^{[1]}}^2{\mu_t^{[2]}}^2[\phi_{12,t} \\
&+ (\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}) \times (\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]})] \\
&+ {\mu_t^{[1]}}^2[\rho_{21}(1+\rho_{21})[Y_{t-1}^{[1]} - \mu_{t-1}^{[1]}] + \rho_{22}(1+\rho_{22})[Y_{t-1}^{[2]} - \mu_{t-1}^{[2]}] + \mu_t^{[2]}(1+\mu_t^{[2]}) - \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) \\
&- \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})] + {\mu_t^{[1]}}^2[\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]}]^2 \\
&- [\phi_{12,t} + (\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}) \\
&\times (\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]})]^2 - 5{\mu_t^{[1]}}^2{\mu_t^{[2]}}^2
\end{aligned} \tag{16}$$

and
$\mathrm{Cov}(Y_t^{[1]}Y_t^{[2]}, Y_{t+h}^{[1]}Y_{t+h}^{[2]} \mid Y_{t-1}^{[1]}, Y_{t-1}^{[2]}, Y_{t+h-1}^{[1]}, Y_{t+h-1}^{[2]})$

$$
\begin{aligned}
&= \phi_{12,t}\phi_{12,t+h} + \mu_{t+h}^{[1]}\mu_{t+h}^{[2]}[\phi_{12,t} + (\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}) \\
&\times (\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]}) + \mu_t^{[2]}\mu_{t+h}^{[2]}[(\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} \\
&- \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}) \times (\rho_{11}Y_{t+h-1}^{[1]} + \rho_{12}Y_{t+h-1}^{[2]} + \mu_{t+h}^{[1]} - \rho_{11}\mu_{t+h-1}^{[1]} - \rho_{12}\mu_{t+h-1}^{[2]})] \\
&+ \mu_{t+h}^{[1]}\mu_t^{[2]}[(\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}) \times (\rho_{21}Y_{t+h-1}^{[1]} + \rho_{22}Y_{t+h-1}^{[2]} + \mu_{t+h}^{[2]} - \rho_{21}\mu_{t+h-1}^{[1]} \\
&- \rho_{22}\mu_{t+h-1}^{[2]})] + \mu_t^{[1]}\mu_{t+h}^{[2]}[(\rho_{11}Y_{t+h-1}^{[1]} + \rho_{12}Y_{t+h-1}^{[2]} + \mu_{t+h}^{[1]} - \rho_{11}\mu_{t+h-1}^{[1]} - \rho_{12}\mu_{t+h-1}^{[2]}) \times (\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} \\
&+ \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]})] + \mu_t^{[1]}\mu_{t+h}^{[1]}[(\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]}) \\
&\times (\rho_{21}Y_{t+h-1}^{[1]} + \rho_{22}Y_{t+h-1}^{[2]} + \mu_{t+h}^{[2]} - \rho_{21}\mu_{t+h-1}^{[1]} - \rho_{22}\mu_{t+h-1}^{[2]})] + \mu_t^{[1]}\mu_t^{[2]}[(\rho_{11}Y_{t+h-1}^{[1]} + \rho_{12}Y_{t+h-1}^{[2]} \\
&+ \mu_{t+h}^{[1]} - \rho_{11}\mu_{t+h-1}^{[1]} - \rho_{12}\mu_{t+h-1}^{[2]}) \times (\rho_{21}Y_{t+h-1}^{[1]} + \rho_{22}Y_{t+h-1}^{[2]} + \mu_{t+h}^{[2]} - \rho_{21}\mu_{t+h-1}^{[1]} - \rho_{22}\mu_{t+h-1}^{[2]})] \\
&- [\phi_{12,t} + (\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}) \\
&\times (\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]})] \\
&\times [\phi_{12,t+h} + (\rho_{11}Y_{t+h-1}^{[1]} + \rho_{12}Y_{t+h-1}^{[2]} + \mu_{t+h}^{[1]} - \rho_{11}\mu_{t+h-1}^{[1]} - \rho_{12}\mu_{t+h-1}^{[2]}) \\
&\times (\rho_{21}Y_{t+h-1}^{[1]} + \rho_{22}Y_{t+h-1}^{[2]} + \mu_{t+h}^{[2]} - \rho_{21}\mu_{t+h-1}^{[1]} - \rho_{22}\mu_{t+h-1}^{[2]})] - 5\mu_t^{[1]}\mu_t^{[2]}\mu_{t+h}^{[1]}\mu_{t+h}^{[2]}
\end{aligned}
\tag{17}
$$

where $\mathrm{Cov}(R_{t+h}^{[1]}, R_{t+h}^{[2]}) = \phi_{12,t+h}$.
$\boldsymbol{D_\psi}$ is a $(T \times 4)$ derivative matrix where

$$
\begin{aligned}
E(Y_t^{[1]}Y_t^{[2]} \mid Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) = &[\kappa_{12,t}[\mu_t^{[1]}(1+\mu_t^{[1]}) - \rho_{11}(1+\rho_{11})\mu_{t-1}^{[1]} - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}(1+\rho_{12})\mu_{t-1}^{[2]} \\
&- \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{\frac{1}{2}}[\mu_t^{[2]}(1+\mu_t^{[2]}) - \rho_{21}(1+\rho_{21})\mu_{t-1}^{[1]} - \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) \\
&- \rho_{22}(1+\rho_{22})\mu_{t-1}^{[2]} - \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{\frac{1}{2}} + (\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} \\
&- \rho_{12}\mu_{t-1}^{[2]}) \times (\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]})]
\end{aligned}
\tag{18}
$$

Thus, the entries of the derivative matrix $\boldsymbol{D_\psi}$ are computed as:

1. $\frac{\partial E(Y_t^{[1]}Y_t^{[2]}|Y_{t-1}^{[1]},Y_{t-1}^{[2]})}{\partial \rho_{11}} = \frac{1}{2}\kappa_{12,t}[\mu_t^{[1]}(1+\mu_t^{[1]}) - \rho_{11}(1+\rho_{11})\mu_{t-1}^{[1]} - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}(1+\rho_{12})\mu_{t-1}^{[2]} - \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{-\frac{1}{2}}[-(1+2\rho_{11})\mu_{t-1}^{[1]} - 2\rho_{11}\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]})][\mu_t^{[2]}(1+\mu_t^{[2]}) - \rho_{21}(1+\rho_{21})\mu_{t-1}^{[1]} - \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{22}(1+\rho_{22})\mu_{t-1}^{[2]} - \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{\frac{1}{2}} + (Y_{t-1}^{[1]} - \mu_{t-1}^{[1]})(\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]})$
2. $\frac{\partial E(Y_t^{[1]}Y_t^{[2]}|Y_{t-1}^{[1]},Y_{t-1}^{[2]})}{\partial \rho_{12}} = \frac{1}{2}\kappa_{12,t}[\mu_t^{[1]}(1+\mu_t^{[1]}) - \rho_{11}(1+\rho_{11})\mu_{t-1}^{[1]} - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}(1+\rho_{12})\mu_{t-1}^{[2]} - \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{-\frac{1}{2}}[-(1+2\rho_{12})\mu_{t-1}^{[2]} - 2\rho_{12}\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})][\mu_t^{[2]}(1+\mu_t^{[2]}) - \rho_{21}(1+\rho_{21})\mu_{t-1}^{[1]} - \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{22}(1+\rho_{22})\mu_{t-1}^{[2]} -$

$\rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{\frac{1}{2}} + (Y_{t-1}^{[2]} - \mu_{t-1}^{[2]})(\rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]})$

3. $\frac{\partial E(Y_t^{[1]}Y_t^{[2]}|Y_{t-1}^{[1]},Y_{t-1}^{[2]})}{\partial \rho_{21}} = \frac{1}{2}\kappa_{12,t}[\mu_t^{[1]}(1+\mu_t^{[1]}) - \rho_{11}(1+\rho_{11})\mu_{t-1}^{[1]} - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}(1+\rho_{12})\mu_{t-1}^{[2]} - \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{\frac{1}{2}}[-(1+2\rho_{21})\mu_{t-1}^{[1]} - 2\rho_{21}\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]})][\mu_t^{[2]}(1+\mu_t^{[2]}) - \rho_{21}(1+\rho_{21})\mu_{t-1}^{[1]} - \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{22}(1+\rho_{22})\mu_{t-1}^{[2]} - \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{-\frac{1}{2}} + (Y_{t-1}^{[1]} - \mu_{t-1}^{[1]})(\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]})$

4. $\frac{\partial E(Y_t^{[1]}Y_t^{[2]}|Y_{t-1}^{[1]},Y_{t-1}^{[2]})}{\partial \rho_{22}} = \frac{1}{2}\kappa_{12,t}[\mu_t^{[1]}(1+\mu_t^{[1]}) - \rho_{11}(1+\rho_{11})\mu_{t-1}^{[1]} - \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}(1+\rho_{12})\mu_{t-1}^{[2]} - \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{\frac{1}{2}}[-(1+2\rho_{22})\mu_{t-1}^{[2]} - 2\rho_{22}\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})][\mu_t^{[2]}(1+\mu_t^{[2]}) - \rho_{21}(1+\rho_{21})\mu_{t-1}^{[1]} - \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{22}(1+\rho_{22})\mu_{t-1}^{[2]} - \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})]^{-\frac{1}{2}} + (Y_{t-1}^{[2]} - \mu_{t-1}^{[2]})(\rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]})$

Thus, the iterative scheme for the second GQL is written as

$$\left(\hat{\boldsymbol{\psi}}_{r+1}\right) = \left(\hat{\boldsymbol{\psi}}_r\right) + [\boldsymbol{D_\psi}'\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\psi}}{}^{-1}\boldsymbol{D_\psi}]_r^{-1}[\boldsymbol{D_\psi}'\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\psi}}{}^{-1}(\boldsymbol{Y_\psi} - \boldsymbol{\mu_\psi})]_r \qquad (19)$$

Using the updated values of $\hat{\boldsymbol{\beta}}$ from equation (14), we solved equation (19) until convergence. $(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})'$ is shown to follow asymptotically the normal distribution with mean 0 and covariance matrix matrix specified by $[\boldsymbol{D_\psi}'\boldsymbol{\Sigma_\psi}{}^{-1}\boldsymbol{D_\psi}]^{-1}$.

4 Simulation Study

This section focuses on simulating BINAR(1) time series data with Geometric marginals under a time-varying covariate design of the form

$$x_{t1} = \begin{cases} -\cos(2\pi t) + 0.01 & (t = 1, \ldots, T/4) \\ \sin(2\pi t) + 0.05 & (t = (T/4) + 1, \ldots, 3T/4) \\ \cos(2\pi t) + 0.10 & (t = (3T/4) + 1, \ldots, T) \end{cases}$$

$$x_{t2} = \begin{cases} (1/t) & (t = 1, \ldots, T/4) \\ (-1/t) & (t = (T/4) + 1, \ldots, 3T/4) \\ t & (t = (3T/4) + 1, \ldots, T) \end{cases}$$

such that $\mu_t^{[k]} = \exp(x_{t1}\beta_1^{[k]} + x_{t2}\beta_2^{[k]})$ and assuming that the inter-relation between $Y_t^{[1]}$ and $Y_t^{[2]}$ is induced by the same innovation series which imply $\kappa_{12,t} = 1$ and $\rho_{12} = \rho_{21} = 0$, $[\rho_{11}, \rho_{22}] = [0.9, 0.9], [0.3, 0.9], [0.3, 0.3]$, $\beta^{[1]} = 0.5$ and $\beta^{[2]} = 0.9$ for $t = 1, 2, \ldots, T = 100, 500, 1000$. We then simulate $R_t^{[k]}$ under the assumed values of $\mu_t^{[k]}$ and (ρ_{11}, ρ_{22}) using the inverse transformation method as in Ristic et al. [14]. 5000 Monte Carlo replications are made for each combination of $[\rho_{11}, \rho_{22}] = [0.9, 0.9], [0.3, 0.9], [0.3, 0.3]$ and the simulated mean estimates are shown below:

ρ_{11}	ρ_{22}	T	Method	$\beta_1^{[1]}$	$\beta_2^{[1]}$	$\beta_1^{[2]}$	$\beta_2^{[2]}$	ρ_{11}	ρ_{22}	ρ_{12}	ρ_{21}	$\kappa_{12,1}$
0.9	0.9	100	GQL	0.4823	0.4870	0.8876	0.8840	0.8819	0.8847	0.0059	0.0047	0.9815
				(0.0910)	(0.0931)	(0.0977)	(0.0946)	(0.1163)	(0.1115)	(0.1125)	(0.1153)	(0.1236)
		500	GQL	0.4923	0.4914	0.8950	0.8917	0.8942	0.8920	0.0017	0.0020	0.9909
				(0.0517)	(0.0512)	(0.0547)	(0.0585)	(0.0649)	(0.0625)	(0.0623)	(0.0630)	(0.0754)
		1000	GQL	0.4991	0.4988	0.8960	0.8981	0.8984	0.8994	0.0004	0.0006	0.9979
				(0.0122)	(0.0171)	(0.0194)	(0.0146)	(0.0271)	(0.0224)	(0.0209)	(0.0237)	(0.0360)
0.3	0.9	100	GQL	0.4894	0.4891	0.8870	0.8840	0.2856	0.8868	0.0099	0.0088	0.9826
				(0.0969)	(0.0977)	(0.0915)	(0.0959)	(0.1133)	(0.1172)	(0.1128)	(0.1119)	(0.1214)
		500	GQL	0.4926	0.4927	0.8944	0.8959	0.2915	0.8928	0.0065	0.0040	0.9918
				(0.0526)	(0.0556)	(0.0507)	(0.0512)	(0.0681)	(0.0694)	(0.0671)	(0.0631)	(0.0728)
		1000	GQL	0.4995	0.4975	0.8969	0.8994	0.2988	0.8963	0.0008	0.0011	0.9995
				(0.0137)	(0.0175)	(0.0111)	(0.0141)	(0.0213)	(0.0293)	(0.0251)	(0.0221)	(0.0332)
0.3	0.3	100	GQL	0.4823	0.4870	0.8804	0.8896	0.2812	0.2834	0.0054	0.0035	0.9819
				(0.0981)	(0.0911)	(0.0931)	(0.0928)	(0.1169)	(0.1160)	(0.1135)	(0.1197)	(0.1274)
		500	GQL	0.4929	0.4942	0.8910	0.8935	0.2964	0.2931	0.0020	0.0015	0.9913
				(0.0594)	(0.0589)	(0.0562)	(0.0580)	(0.0614)	(0.0621)	(0.0677)	(0.0681)	(0.0717)
		1000	GQL	0.4956	0.4992	0.8987	0.8990	0.2988	0.2980	0.0004	0.0001	0.9966
				(0.0152)	(0.0125)	(0.0135)	(0.0133)	(0.0241)	(0.0208)	(0.0219)	(0.0211)	(0.0375)

Table 1. GQL estimates of the parameters and standard errors under non-stationary BINAR(1) process based on 5000 Monte-Carlo replications.

From Table 1, we note that the estimates at the different size points and under the different serial parameters yield consistent estimates. The cross-correlation, $\kappa_{12,1}$ is close to unity. With increasing time points, the standard errors for the different estimates decreases irrespective of low or high serial-correlation. Some details on the number of non-convergent simulations are: For $\rho_{11} = \rho_{22} = 0.9$ under GQL, around 360 simulations failed for $T = 100$, 300 for $T = 500$ and 220 for $T = 1000$. For $\rho_{11} = 0.3$ and $\rho_{22} = 0.9$, around 340 GQL simulations failed for $T = 100$, 275 for $T = 500$ and 190 for $T = 1000$. However, when $\rho_{11} = \rho_{22} = 0.3$, the GQL algorithms flopped in 315 simulations for $T = 100$, 215 for $T = 500$ and 170 for T=1000. The failure were mainly due to either the ill-conditioned covariance matrix (12) or the Hessian structure in equation (14).

5 Conclusion

This paper addresses the modelling of the full non-stationary BINAR(1) process with marginal Geometric counting series. In this process, it is noticed that the joint probability function of the innovation series is unknown and this limits the construction of the conditional likelihood function to ultimately estimate the unknown parameters of the model. Alternatively, the paper proposes the GQL approach that requires only the correct specification of the score function and its corresponding moment vectors. As for the auto-covariance components of the GQL technique, these were derived using the modelling assumptions and the multivariate normality structure. We also note that under the re-parametrization of some coefficients, the proposed BINAR(1) process reduces to the models proposed by Ristic et al. [15] and Nastic et al. [8]. In the numerical experiments, the GQL approach was shown to provide consistent parameter estimates.

References

1. FitzMaurice, G., Laird, N., Rotnizky, A.: Regression models for discrete longitudinal responses. Statistical Science 8, 284–309 (1993)

2. Jowaheer, V., Mamode Khan, N., Sunecher, Y.: A non-stationary BINAR(1) process with negative binomial innovations for modeling the number of goals in the first and second half: the case study of Arsenal Football Club. Communication in Statistics-Case Studies, Data Analysis and Applications (2017a)
3. Jowaheer, V., Sutradhar, B.: Analysing longitudinal count data with overdipsersion. Biometrika 89, 389–399 (2002)
4. Mamode Khan, N., Jowaheer, V.: Comparing joint GQL estimation and GMM adaptive estimation in COM-Poisson longitudinal regression model. Commun Stat-Simul C. 42(4), 755–770 (2013)
5. Mamode Khan, N., Sunecher, Y., Jowaheer, V.: Modelling a non-stationary BINAR(1) Poisson process. Journal of Statistical Computation and Simulation 86, 3106–3126 (2016b)
6. Mamode Khan, N., Sunecher, Y., Jowaheer, V.: Modelling non-stationary INMA(1) time series of counts with COM-poisson innovations (2016d)
7. Nastic, A., Laketa, P., Ristic, M.: Random environment integer-valued autoregressive process. Journal of Time Series Analysis 37(2), 267–287 (2016a)
8. Nastic, A., Ristic, M., Popovic, P.: Estimation in a bivariate integer-valued autoregressive process. Communication in Statistics-Theory and Methods 45(19), 5660–5678 (2016b)
9. Pedeli, X., Karlis, D.: Bivariate INAR(1) models. Tech. rep., Athens University of Economics (2009)
10. Pedeli, X., Karlis, D.: A bivariate INAR(1) process with application. Statistical Modelling: An International Journal 11, 325–349 (2011)
11. Pedeli, X., Karlis, D.: Some properties of multivariate INAR(1) processes. Computational Statistics and Data Analysis 67, 213–225 (2013a)
12. Pedeli, X., Karlis, D.: On composite likelihood estimation of a multivariate INAR(1) model. Journal of Time Series Analysis 34, 206–220 (2013c)
13. Prentice, R., Zhao, L.: Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 47, 825–39 (1991)
14. Ristic, M., Bakouch, H., Nastic, A.: A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. Journal of Statistical Planning and Inference 136, 2218–2226 (2009)
15. Ristic, M., Nastic, A., Jayakumar, K., Bakouch, H.: A bivariate INAR(1) time series model with geometric marginals. Applied Mathematical Letters 25(3), 481–485 (2012)
16. Silva, M., Oliveira, V.: Difference equations for the higher order moments and cumulants of the INAR(1) model. Journal of Time Series Analysis 25, 317–333 (2004)
17. Steutel, F., Van Harn, K.: Discrete analogues of self-decomposability and stability. The Annals of Probability 7, 3893–3899 (1979)
18. Steutel, F., Van Harn, K.: Discrete operator self-decomposability and queing networks. Stochastic Models 2, 161–169 (1986)
19. Sunecher, Y., Mamodekhan, N., Jowaheer, V.: A gql estimation approach for analysing non-stationary over-dispersed BINAR(1) time series. Journal of Statistical Computation and Simulation (2017)
20. Sutradhar, B.: An overview on regression models for discrete longitudinal responses. Statistical Science 18(3), 377–393 (2003)
21. Sutradhar, B., Jowaheer, V., Rao, P.: Remarks on asymptotic efficient estimation for regression effects in stationary and non-stationary models for panel count data. Brazilian Journal of Probability and Statistics 28(2), 241–254 (2014)

22. Weiβ, C.: Thinning operations for modelling time series of counts-a survey. AStA Advances in Statistical Analysis 92(3), 319–341 (2008b)

A

The conditional expectations, variances, covariances:

1. $\mathrm{E}(Y_t^{[1]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) = \rho_{11}Y_{t-1}^{[1]} + \rho_{12}Y_{t-1}^{[2]} + \mu_t^{[1]} - \rho_{11}\mu_{t-1}^{[1]} - \rho_{12}\mu_{t-1}^{[2]}$
2. $\mathrm{E}(Y_t^{[2]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) = \rho_{21}Y_{t-1}^{[1]} + \rho_{22}Y_{t-1}^{[2]} + \mu_t^{[2]} - \rho_{21}\mu_{t-1}^{[1]} - \rho_{22}\mu_{t-1}^{[2]}$
3. $\mathrm{Var}(Y_t^{[1]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) = \rho_{11}(1+\rho_{11})[Y_{t-1}^{[1]} - \mu_{t-1}^{[1]}] + \rho_{12}(1+\rho_{12})[Y_{t-1}^{[2]} - \mu_{t-1}^{[2]}] + \mu_t^{[1]}(1+\mu_t^{[1]})$
   $- \rho_{11}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{12}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})$
4. $\mathrm{Var}(Y_t^{[2]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) = \rho_{21}(1+\rho_{21})[Y_{t-1}^{[1]} - \mu_{t-1}^{[1]}] + \rho_{22}(1+\rho_{22})[Y_{t-1}^{[2]} - \mu_{t-1}^{[2]}] + \mu_t^{[2]}(1+\mu_t^{[2]})$
   $- \rho_{21}^2\mu_{t-1}^{[1]}(1+\mu_{t-1}^{[1]}) - \rho_{22}^2\mu_{t-1}^{[2]}(1+\mu_{t-1}^{[2]})$
5. $\mathrm{Cov}(Y_t^{[k]}, Y_{t+h}^{[k]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}, Y_{t+h-1}^{[1]}, Y_{t+h-1}^{[2]}) = 0, \quad \text{for} \quad h \neq 0$
6. $\mathrm{Cov}(Y_t^{[1]}, Y_t^{[2]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) = \mathrm{Cov}(R_t^{[1]}, R_t^{[2]})$
7. $\mathrm{Cov}(Y_t^{[1]}Y_{t+h}^{[2]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}, Y_{t+h-1}^{[1]}, Y_{t+h-1}^{[2]}) = 0, \quad \text{for} \quad h \neq 0$
8. $\mathrm{Cov}(Y_{t+h}^{[1]}Y_t^{[2]}|Y_{t-1}^{[1]}, Y_{t-1}^{[2]}, Y_{t+h-1}^{[1]}, Y_{t+h-1}^{[2]}) = 0, \quad \text{for} \quad h \neq 0$

Characterising Dependency in Computer Networks Using Spectral Coherence

Alex J. Gibberd, Jordan Noble, and Edward A.K. Cohen*

Department of Mathematics, Imperial College London
South Kensington Campus, London, SW7 2AZ, United Kingdom
agibberd@ic.ac.uk,jnoble@ic.ac.uk,ecohen@ic.ac.uk

Abstract. The quantification of normal and anomalous traffic flows across computer networks is a topic of pervasive interest in network security, and requires the timely application of time-series methods. The transmission or reception of packets passing between computers can be represented in terms of time-stamped events and the resulting activity understood in terms of point-processes. Interestingly, in the disparate domain of neuroscience, models for describing dependent point-processes are well developed. In particular, spectral methods which decompose second-order dependency across different frequencies allow for a rich characterisation of point-processes. In this paper, we investigate using the spectral coherence statistic to characterise computer network activity, and determine if, and how, device messaging may be dependent. We demonstrate on real data, that for many devices there appears to be very little dependency between device messaging channels. However, when significant coherence is detected it appears highly structured, a result which suggests coherence may prove useful for discriminating between types of activity at the network level.

1 Introduction

Understanding how devices on computer networks communicate is a challenging task. While it is possible to gather vast quantities of data from such networks, for instance via packet monitoring, it is difficult to store, let alone process. As a result, protocols such as NetFlow which sample and summarise packet level data are now very popular (Duffield, 2004). Even still, regular monitoring protocols can produce hundreds of gigabytes of summary statistics on a network per day which need to be converted into actionable insights for network administrators. In the context of anomaly detection, one may be interested in either detecting anomalies online, or in a forensic context.

Network defenders should be at a theoretical advantage over attackers, in that they can attempt to model and understand the day-to-day activity of their network. From such models they can then define what anomalous, and/or malicious events may look like. Additionally, in order to enhance detection performance,

* This work was funded as part of EPSRC grant EP/P011535/1

one may desire to use prior knowledge of what benign network activity should look like in order to define anomalies. For example, and relevant to the approach developed here, one may expect that communication between pairs of devices are not correlated such that their activity should be broadly independent when monitored at a network level. However, if traffic is dependent across pairs then this may indicate potentially malicious behaviour such as lateral movement or tunnelling (Neil et al., 2013).

As summarised in Ahmed et al. (2016), anomaly detection methods may generally be broken down into classification, statistical, or clustering based approaches. However, in any of these guises, anomaly detection will fail if the features extracted from the data fail to appropriately quantify network behaviour. In the context of modelling dependency between data streams, there are a great variety of measures and methods that can be used, for instance through measures such as covariance (Jin et al., 2007), correlation (Neil et al., 2013), partial correlation (Gibberd et al., 2017), or higher-order measures such as cross-cumulants (Brillinger, 1993). A traditional approach to network traffic modelling is to assume it is generated according to a Poisson point process (Duffield, 2004; Murdoch and Zieliński, 2007). While such models may be generalised to a multivariate setting (Baurele and Grubel, 2005), they do not allow for us to encode auto-correlation structure within a point-process. As a result, we may be able to describe processes which are dependent on each other across data-streams, but they do not allow for dependency within a data-stream itself. When considering computer network traffic, it is not hard to imagine that events from a device will be somehow dependent on previous events from that same device. Spectral approaches, based on either Fourier (Brillinger, 1981) or time-scale wavelet analysis (Riedi et al., 1999; Scherrer and Larrieu, 2007) of processes provide a valuable tool in this situation as they allow for both a rich description of auto-correlation and cross-channel dependency (Cohen, 2014).

In this paper, we propose to utilise a measure known as spectral coherence (Carter, 1987) to characterise dependency between network communication channels. We are not aware of any previous application of such a measure to network traffic analysis, although the method has received great attention in neuroscience for modelling neuron dependency (Jarvis and Mitra, 2001).

2 Dataset and Preprocessing

Consider network connected devices A, B, C and their associated users, for example, these may be personal computers, DNS servers, authentication servers, or even printers. Typically, we would expect these devices to go about their work as fairly independent actors, i.e. they may browse websites, download material etc., but not in any particularly coordinated manner. Device communication is typically performed through packet transmission. However, given network monitoring limitations, the events that we analyse need not necessarily be packets themselves. More likely, they are aggregates or summaries of communication, for example NetFlow sessions.

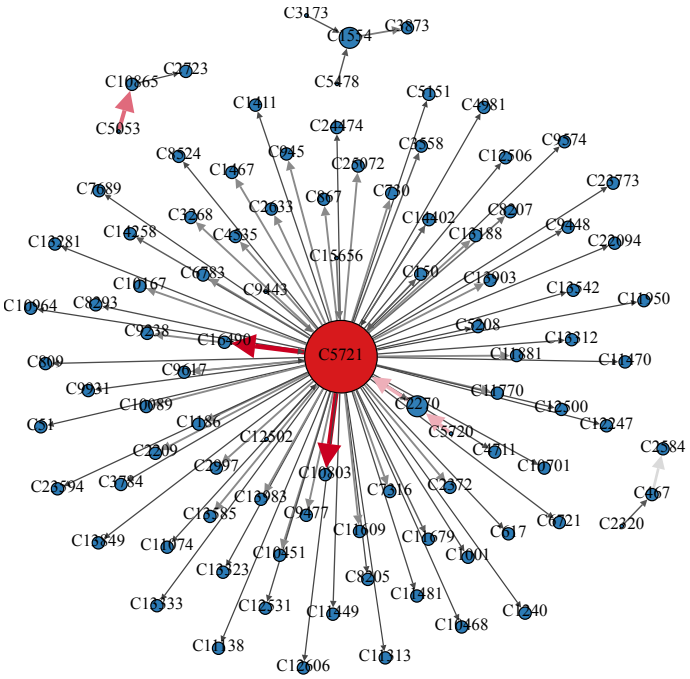
In our case, we analyse a subset of NetFlow session data from the Los Alamos National Laboratory (LANL) multi-source cyber-security events data (Kent, 2015a,b). More specifically, we create a subset of events (NetFlow session start times) relating to the top $N_{\text{triple}} = 500$ busiest edge-pairs in the network over a single day's (Thursday) worth of data. We assess dependency in a pair-wise manner such that data-streams correspond to directed edge-communication between devices $A \to B$ and $B \to C$ for $i = 1, \ldots, N_{\text{triple}}$ device triples $(A, B, C)_i$. For each triple, the activity for each edge corresponds to the same time frame. The protocol monitored is the same for all edge pairs. To avoid confusion, we exclude all flows from the triple (C, B, A) when the triple (A, B, C) is included. One should note that our selection criteria for data-set construction does not explicitly specify devices which have a particular function on the network. However, if we look at the graph of communication edges in Figure 1 it appears that many of our triples have repetitive edges, there are only 95 unique devices in our data-set and 96 unique edges. Looking at the topology of the network it appears that most of the devices are communicating through the device C5721, while we do not have labelled data relating to the function of devices, it would appear that this node acts as some form of server.

We note that in our recordings it is possible to observe two events which have the same start time. This means that the events cannot reasonably be treated as being observed in continuous time, and indeed the timestamps provided with our data are only accurate to the second. As such, the raw events are aggregated into bins of width $\Delta = 1$s. The binned bivariate process will be denoted $\{\boldsymbol{X}[k] = [X_{AB}[k], X_{BC}[k]]^T; k \in \mathbb{Z}\}$, for which we observe a portion $\boldsymbol{X}[1], \boldsymbol{X}[2], ..., \boldsymbol{X}[K]$. As a pre-processing step, we subtract the empirical mean of the data-streams, relabeling $\boldsymbol{X}[k] := \boldsymbol{X}[k] - \bar{\boldsymbol{X}}$ so that they can be well approximated as zero-mean processes.

3 Spectral Coherence as a Measure of Association

In this section, we will define the spectral coherence as a property relating to the cross-spectrum of a bivariate process. The discussion here will be developed based on the understanding that we are with observations relating to a discrete-time process. However, it is also possible to perform such analysis at the individual event level, for examples, see the work of Cohen (2014); Jarvis and Mitra (2001); Brillinger (1972).

To simplify the notation, let us use $X_1[k] \equiv X_{AB}[k]$ and $X_2[k] \equiv X_{BC}[k]$. Furthermore, we will assume that $\{X_1[k], X_2[k]\}$ represent a jointly second-order stationary process, i.e. the covariance $\Sigma_{ij}[\tau] \equiv \text{Cov}(X_i[k+\tau], X_j[k])$ only depends on τ for $i, j = 1, 2$. Provided $\sum_{\mathcal{T}} |\Sigma_{ij}[\tau]| < \infty$, then for all frequencies $|\omega| \leq \pi/\Delta$, the function $S_{ij}(\omega) = \Delta \sum_{\tau=-\infty}^{\infty} \Sigma_{ij}[\tau] e^{-i\omega\tau\Delta}$, $i, j = 1, 2$, is termed the spectum of $\{X_i[k]\}$ when $i = j$, and the *cross-spectrum* between $\{X_1[k]\}$ and $\{X_2[k]\}$ when $i \neq j$. These spectra can be conveniently represented with the spectral matrix $\boldsymbol{S}(\omega) = (S_{ij}(\omega))$. The argument we present in this paper, is that the cross-spectrum provides a rich framework within which we may characterise



N_{triple}	500	Average length T	17.2 (hours)
Number of nodes	95	Average rate λ_{AB}	0.101 (1/s)
Unique edges	96	Average rate λ_{BC}	0.032 (1/s)

Fig. 1. Graph of messaging channels under analysis. Size of text and node are respectively proportional to the out-going and in-going degree. The weight of the edges (and colour) is proportional to the rate measured on that edge.

computer network messaging processes. In particular, since we are interested in dependency between data-streams, we will concern ourselves with the squared coherencey, or ordinary coherence, defined as the real-valued quantity

$$R(\omega) = \frac{|S_{12}(\omega)|^2}{S_{11}(\omega)S_{22}(\omega)} \, . \tag{1}$$

The coherence provides a useful statistic for assessing dependency between point-processes; not only does it permit a decomposition over frequencies allowing one to highlight periodicities associated with dependence, but is also invariant to scaling of the marginal auto-covariance as the measure is normalised by the on-diagonal spectra.

3.1 Estimating the Spectra of Point-Processes

Since the true values of the spectra, cross-spectra and coherence are unknown to us, we are required to estimate them from data. The approach that we utilise here is based on the work of Thomson (1982). Specifically, we will construct our estimators from the *tapered discrete-time Fourier transform (tDFT)* defined as

$$\hat{F}_{j;l}(\omega) \equiv \Delta^{1/2} \sum_{k=1}^{K} h_l[k] X_j[k] e^{-i\omega k\Delta} \quad j = 1, 2, \tag{2}$$

for frequencies $-\pi/\Delta < \omega < \pi/\Delta$ where $\{h_l[k]; k = 1, ..., N\}$ for $l = 1, \ldots, L$ are a set of taper sequences. The tapers in the above construction are important, in that they enable us to selectively transform data-points prior to taking the Fourier transform.

If we temporarily assume that $h_l[k] = 1$ for all l, k, then taking the conjugate outer-product leads to the periodogram $\hat{I}_{ij;l}(\omega) \equiv \hat{F}_{i;l}(\omega)\hat{F}^*_{j;l}(\omega)$. Unfortunately, while the periodogram is an asymptotically unbiased estimator of the spectrum $E[\hat{I}_{ij;l}(\omega)] \to S_{ij}(\omega)$ as $T \to \infty$, it is not consistent, in that $\mathrm{Var}[\hat{I}_{ij;l}(\omega)] \not\to 0$. Principally, this is due to us attempting to estimate the spectra at an infinite number of frequencies $\omega \in \mathbb{R}$ with only a finite portion of data (Brillinger, 1972; Jarvis and Mitra, 2001).

There are several approaches which can be used to sculpt asymptotically consistent estimators of the spectra (Nuttall and Carter, 1982; Thomson, 1982; Walden, 2000). A general strategy (Walden, 2000), is to adapt the direct spectral estimate (where $h_l[k] = 1$) such that the tapers take different shapes, for example they may be supported in disjoint regions (Brillinger, 1981), or constitute a set of overlapping windows (Nuttall and Carter, 1982). From the Fourier transform of the tapered data, we may then construct vectors $\hat{\boldsymbol{F}}_l(\omega) = [\hat{F}_{1;l}, \hat{F}_{2;l}]^T$ and compute what is known as a *multi-taper spectral estimator* by averaging:

$$\hat{\boldsymbol{S}}(\omega) = \frac{1}{L} \sum_{l=1}^{L} \hat{\boldsymbol{F}}_l(\omega) \hat{\boldsymbol{F}}_l^H(\omega) \quad i, j = 1, 2,$$

where ^{H} denotes the complex conjugate transpose. From the multi-taper spectral estimate we may then obtain an estimate for the coherence $\hat{R}_{12}(\omega)$ via (1) replacing the true spectra $S_{11}(\omega)$, $S_{22}(\omega)$ and $S_{12}(\omega)$ with the estimates $\hat{S}_{11}(\omega)$, $\hat{S}_{22}(\omega)$ and $\hat{S}_{12}(\omega)$, respectively.

3.2 Taper Specification

If we consider choosing orthogonal taper sequences whereby $\sum_k h_l[k]h_{l'}[k] = 0$ for $l \neq l'$ then the resultant Fourier transforms will be asymptotically independent (Brillinger, 1981). Averaging over these independent sequences can then reduce the variation in the estimate, the reduction will be related to the number of tapers we average over. It has been demonstrated, c.f. Brillinger (1981), that in the case of asymptotically orthogonal tapers, the sampling distribution of the

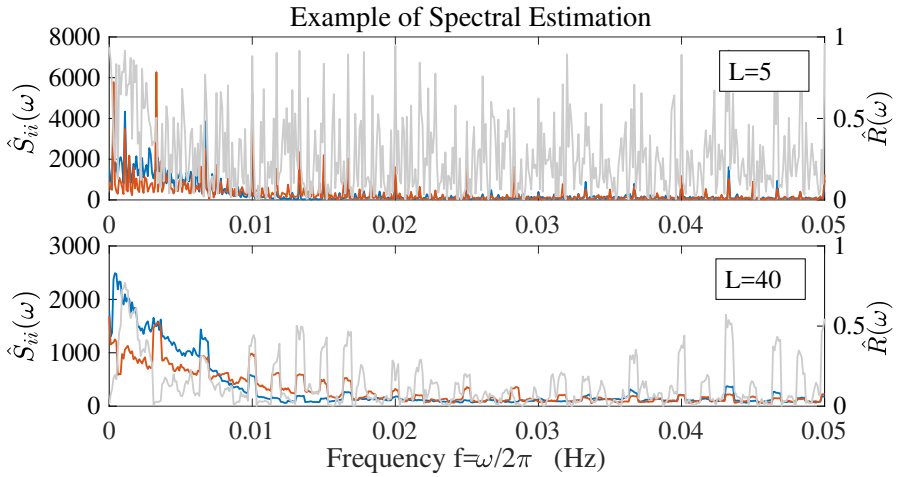


Fig. 2. Example of estimates for the spectral density. Top: estimation with $L = 5$ tapers. Bottom: estimation with $L = 40$ tapers. The red and blue lines respectively illustrate the on-diagonal spectral density $\hat{S}_{AB,AB}(\omega)$ and $\hat{S}_{BC,BC}(\omega)$. The grey line indicates the resultant coherence $\hat{R}(\omega)$.

spectral matrix $\hat{\boldsymbol{S}}(\omega) = (\hat{S}_{ij}(\omega))$ is given by a 2D complex Wishart distribution $\hat{\boldsymbol{S}}(\omega) \sim W_2^C(L, \boldsymbol{S}(\omega))$ with L degrees of freedom and scale matrix $\boldsymbol{S}(\omega)$.

In our application, we utilise a form of taper first demonstrated for spectral estimation by Thomson (1982). Often referred to as the Slepian tapers, these sequences have the beneficial properties that they are mutually orthogonal while maximising energy concentration in a small frequency interval $[-\omega_W, \omega_W]$. If two frequencies are separated by more than this bandwidth, then the bias due to tapering is in some sense minimised. However, as the number of tapers L increases, the width of the side-lobe associated with the Fourier transform of $h_L[k]$ necessarily increases. As such, there is a classic bias-variance trade-off, increasing L reduces the variance, but increases bias. The appropriate number of tapers to use is highly dependent on application and something we will shortly revisit in the context of the network traffic dataset.

4 Dependency in Network Traffic

Applying coherence estimation to edge pairs results in a set of estimates

$$\{\hat{R}^{(n)}(\omega_1, \ldots, \omega_{N_f})\} \quad \text{for } n = 1, \ldots, N_{\text{triple}} \, .$$

As may be expected there is significant variation of the spectra across the set of edge-pairs. In this analysis, we consider fixing the window of frequencies such that $\omega_q = 2\pi f_{\max}(q/N_f)$ for $q = 1, \ldots, N_f = 500$ and $f_{\max} = 0.05Hz$. As the length of the edge-pair recordings differ, one may desire to increase L as

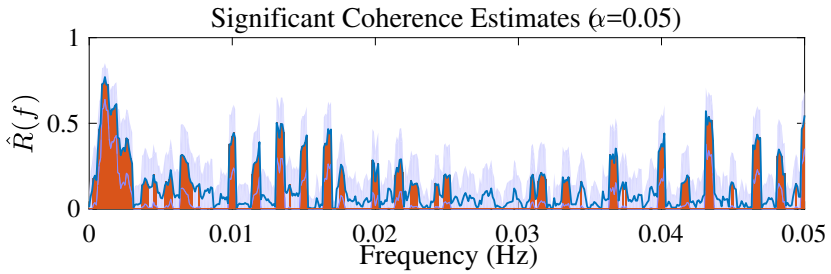


Fig. 3. The coherence as plotted in the bottom of Fig. 2 with two-sided 95% confidence intervals. Frequencies where the confidence interval excludes zero are highlighted in red.

a function of length T. Potentially, this would lead to increased confidence in our spectral estimate as the Wishart degrees of freedom are increased. However, such an adaptive tapering scheme where L depends on T creates challenges when comparing across coherence estimates as it may be hard to disentangle differences due to the tapering treatment from underlying differences in the process spectra. As such, in these experiments we decide to fix the number of tapers at a moderate level $L = 40$ for all edge-pairs. The difference between tapering with $L = 5$ and $L = 40$ is demonstrated in Figure 2. Note, that while the cross-spectra for different edge-pairs may be of a different scale, the coherence (plotted in grey) provides an intuitive measure on $[0, 1]$ allowing comparison across many data-stream pairs. As an aside, the individual spectra appear non-Poisson, exhibiting shapes that are characteristic of self-exciting behaviour (Hawkes, 1971).

Acknowledging that there will be some error in our spectral estimates, it is desirable to assign some measure of confidence to estimates. A useful corollary of the Wishart asymptotic result for multi-taper estimates is that the coherence is distributed (asymptotically) according to the Goodman distribution (Carter, 1987; Goodman, 1963). Based on this distribution, there are a variety of tests that one may perform to assess the *significance* of a coherence estimate. For example, one may test the null hypothesis that states $R(\omega_q) = 0$ for each frequency ω_q, $q = 1, \ldots, N_f$. Rather than test explicitly against a null of zero coherence, in this work we construct two sided confidence intervals in a similar manner to Wang and Tang (2004). Examples of such intervals for $\alpha = 0.05$ are reported in Fig. 3. Alongside these intervals denoted $[\hat{a}_{\alpha/2}, \hat{a}_{1-\alpha/2}]$, we declare that the coherence at a frequency is significant if the interval excludes zero. For this particular triple, we note what appears to be significant coherent beaconing across the devices at multiples of approx. 0.0017Hz (a periodicity of 10 minutes).

To assess variation in coherence estimates across the $N_{\text{triple}} = 500$ edge-pairs under study we attempt to cluster the resultant coherence estimates. Prior to performing clustering, we threshold coherence estimates according to the confidence intervals such that $\hat{R}^*(\omega_q) = 0$ if $0 \in [\hat{a}_{\alpha/2}, \hat{a}_{1-\alpha/2}]$ and $\hat{R}^*(\omega_q) = \hat{R}(\omega_q)$ otherwise. The resultant coherence estimates are then modelled as a *Gaussian*

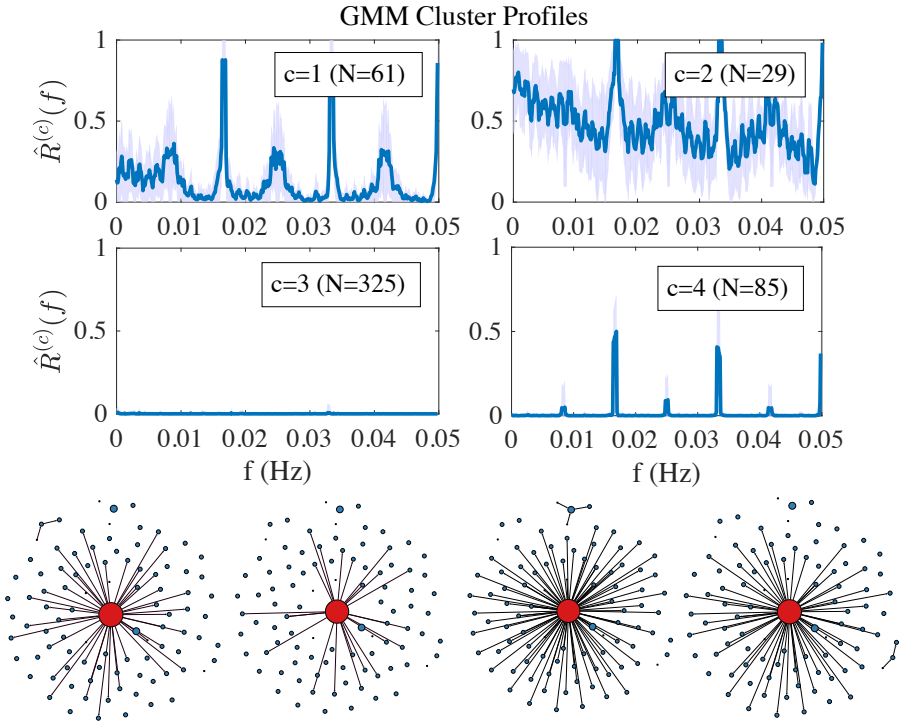


Fig. 4. GMM clustering of the significance thresholded coherence estimates with $C = 4$. Shaded regions indicate points within one standard-deviation from the cluster mean. Bottom: graph of edges relating to each cluster $c = 1, 2, 3, 4$ from left to right.

mixture model (GMM), such that $[\hat{R}^*(\omega_1), \ldots, \hat{R}^*(\omega_{N_f})] \sim \text{GMM}(\{\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c\}_{c=1}^C)$ where $\boldsymbol{\mu}_c \in \mathbb{R}^{N_f}$ represent cluster means and $\boldsymbol{\Sigma}_c$ the cluster covariances.

While the Gaussian assumption of the above model contrasts with the Goodman asymptotic distribution for the coherence, the approximation may still hold relevance. For instance, Enochson and Goodman (1965) demonstrated that a Gaussian approximation may be effective when calculating confidence intervals. In this example we use the MATLAB implementation of the expectation-maximisation with covariance regularisation set at $\lambda = 0.001$. Due to the many local-minima that may be obtained when fitting a GMM, we perform one thousand replications with random initialisation and report the clustering with the largest likelihood. Figure 4 presents the resulting mean profiles of $C = 4$ clusters alongside the standard-deviation obtained from the estimated covariance matrices.

5 Discussion

The clustering results are insightful in that the emergence of clusters $c = 3, 4$ partially confirm our initial hypothesis that many device pairs exhibit little dependency. Out of the initial 500 edge-pairs 90 are placed into clusters with non-negliable coherence, for reference, the example demonstrated in Fig. 2 is placed into cluster $c = 1$. These active coherent clusters exhibit pronounced structure across multiple frequencies, once again providing evidence that modelling auto-covariance within network traffic is important. Of some note, is the clear peak at $f = 0.018Hz$ corresponding to periodicity of around 57s . Without a more intimate knowledge of device functionality on the network we can only hypothesise the cause of such a feature, but it is possible this is due to beaconing activities. Interestingly, Heard et al. (2014) demonstrate strong periodicity at this same frequency for devices connecting with dropbox.com.

The results obtained here suggest that coherence may be considered a valuable feature when analysing dependency in network traffic flows. For instance, while previous methods of analysis (Gibberd et al., 2017; Ahmed et al., 2016) rely on detecting direct correlation between data-streams, they do not allow for us to decompose this dependency over frequency dimension. Our analysis here, demonstrates that in practice there is much structure in this frequency dimension, for example, as seen in Fig. 4. We posit that most of this structure is actually due to benign beaconing activity, for example, synchronising with external services. However, scanning techniques used by intruders may also create similar strongly periodic activity and it is thus of interest to administrators to detect such patterns. The coherence feature should thus be of significant interest to those developing forensic, or online anomaly detection systems.

As a direction for future work, it is possible the methods developed here to detect association between event driven data-streams in a defensive context, may also be used as a form of correlation attack, for example to break anonymisation protocols when traffic is transmitted via mixing devices (Murdoch and Zieliński, 2007). Alternatively, one may investigate the use of the coherence statistic in other forms of anomaly detection methodology, for instance to enhance the clustering and classification approaches described in Ahmed et al. (2016). On a methodological level, it may also be of interest to relax the stationarity assumptions of this work, for instance within a wavelet framework. Indeed an algorithm that could derive the coherence in a streaming manner would be an important step towards building a practical anomaly detector.

Bibliography

Ahmed, M., Naser Mahmood, A., Hu, J.: A survey of network anomaly detection techniques. Journal of Network and Computer Applications 60, 19–31 (2016)

Baurele, N., Grubel, R.: Multivariate Counting Processes: Copulas and Beyond. Astin Bulletin 35(2), 379–408 (2005)

Brillinger, D.R.: The spectral analysis of stationary interval functions. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics pp. 483–513 (1972)

Brillinger, D.R.: Time Series: Data Analysis and Theory. SIAM, Philadelphia (1981)

Brillinger, D.: Moments, Cumulants, and Some Applications to Stationary Random Processes. Tech. rep., University College Berkley (1993)

Carter, G.: Coherence and time delay estimation. Proceedings of the IEEE 75(2), 236–255 (1987)

Cohen, E.A.K.: Multi-wavelet coherence for point processes on the real line. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (2), 2649–2653 (2014)

Duffield, N.: Sampling for Passive Internet Measurement: A Review. Statistical Science 19(3), 472–498 (2004)

Enochson, L., Goodman, N.: Gaussian approximations to the distribution of sample coherence. Technical Report (June) (1965)

Gibberd, A., Evangelou, M., Nelson, J.D.: The Time-Varying Dependency Patterns of NetFlow Statistics. In: IEEE International Conference on Data Mining Workshops, ICDMW. pp. 288–294 (2017)

Goodman, N.R.: Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction). The Annals of Mathematical Statistics 34(1), 152–177 (1963)

Hawkes, A.G.: Spectra of Some Self-Exciting and Mutually Exciting Point Processes. Biometrika 58(1), 83–90 (1971)

Heard, N., Rubin-Delanchy, P., Lawson, D.J.: Filtering automated polling traffic in computer network flow data. Proceedings - 2014 IEEE Joint Intelligence and Security Informatics Conference, JISIC 2014 pp. 268–271 (2014)

Jarvis, M.R., Mitra, P.: Sampling Properties of the Spectrum and Coherency of Sequences of Action Potentials. Neural Computation 749, 717–749 (2001)

Jin, S., Yeung, D.S., Wang, X.: Network intrusion detection in covariance feature space. Pattern Recognition 40(8), 2185–2197 (2007)

Kent, A.D.: Comprehensive, Multi-Source Cyber-Security Events. Los Alamos National Laboratory (2015a)

Kent, A.D.: Cybersecurity Data Sources for Dynamic Network Research. In: Dynamic Networks in Cybersecurity. Imperial College Press (Jun 2015b)

Murdoch, S.J., Zieliński, P.: Sampled Traffic Analysis by Internet-Exchange-Level Adversaries. Privacy Enhancing Technologies pp. 167 – 183 (2007)

Neil, J., Hash, C., Brugh, A., Fisk, M., Storlie, C.C.B.: Scan Statistics for the Online Detection of Locally Anomalous Subgraphs. Technometrics 55(4), 403–414 (2013)

Nuttall, A., Carter, G.: Spectral estimation using combined time and lag weighting. Proceedings of the IEEE 70(9), 1115–1125 (1982)

Riedi, R.H., Crouse, M.S., Ribeiro, V.J., Baraniuk, R.G.: A multifractal wavelet model with application to network traffic. IEEE Transactions on Information Theory 45(3), 992–1018 (1999)

Scherrer, A., Larrieu, N.: Non-gaussian and long memory statistical characterizations for internet traffic with anomalies. Dependable and ... 4(1), 56–70 (2007)

Thomson, D.J.: Spectrum Estimation and Harmonic Analysis. Proceedings of the IEEE 70(9) (1982)

Walden, A.T.: A unified view of multitaper multivariate spectral estimation. Biometrika 87(4), 767–788 (2000)

Wang, S., Tang, M.: Exact confidence interval for magnitude-squared coherence estimates. Signal Processing Letters, IEEE 11(3), 326 (2004)

Keyword Index

.

Author Index