8th International Conference on Time Series and Forecasting

ITISE-2022

PROCEEDINGS OF ABSTRACTS

27th-30th JUNE, 2022
Gran Canaria (SPAIN)

ITISE 2022
International Conference on Time Series and Forecasting

Proceedings of abstracts 27-30 June 2022 Gran Canaria (Spain)

Editors and Chairs
Olga Valenzuela
Fernando Rojas
Héctor Pomares
Luís Javier Herrera
Ignacio Rojas

I.S.B.N: 978-84-19214-24-9
Legal Deposit: GR 954-2022
Edit and Print: Godel Impresiones Digitales S.L.

All rights reserved to authors. The total or partial reproduction of this work is strictly prohibited, without the strict authorization of the copyright owners, under the sanctions established in the laws.

ITISE-2022 Program

Monday, June 27th, 2022	
18:30-20:00	**REGISTRATION DESK** ***(start at 18:30h but it is open during all the conference)***
18:30-20:00	**Upload the presentations to the room's computer (in case you haven't sent them by email).**

NOTES:

- All **Sessions A** will be held in Hotel Lopesan Villa del Conde Resort. They are **face-to-face sessions**, and they will also be shared on-line by Zoom. The **plenary lectures** are in **Session A.**

- All **Sessions B** will be held on-line (virtual) using Zoom.

- Oral Presentation: 12-14 minutes. Short Presentation: 7 minutes.

Session A: Located in the last floor of the main building

Tuesday, June 28, 2022		
9:00	**REGISTRATION DESK** ***(start at 9:00h but it is open during all the conference)***	
	All **Sessions A: Oral face-to-face sessions**. All **Sessions B:** Oral (will be held on-line by Zoom)	
9:15-10:15	**Session A.1: Time Series Analysis with Computational Intelligence (part I)**	
10:20- 11:20	**Session A.2: Advanced Econometric Methods**	**Session B.1: Applications in: Energy, Finance, Transportation, Networks, Meteorology, Health, Research and Environment (part I)**
11:20-12:00	**COFFEE BREAK**	
12:00-12:55	**Session A.3: Time Series Analysis of Cryptocurrency (High-Frequency)**	**Session B.2: Nonstationarity**
13:00-14:00	**Session A.P1: Opening & Plenary Lecture.** **Prof. Daniel Peña Sánchez De Rivera** **University Carlos III, Madrid (Spain)**	
14:00-16:00	**REST BREAK**	
16:00-17:00	**Session A.4: Applications in: Energy, Finance, Transportation, Networks, Meteorology, Health, Research and Environment, etc**	**Session B.3: Advances in Time Series Analysis and Forecasting. Short Presentation (part I)**
17:05-18:20	**Session A.5: Energy forecasting**	
18:20-19:30	**Session A.6: POSTER SESSION**	

Wednesday, June 29th, 2022		
9:00	**REGISTRATION DESK** ***(start at 9:00h but it is opened during all the conference)***	
	All **Sessions A: Oral face-to-face sessions**. All **Sessions B:** Oral (will be held on-line by Zoom)	
9:00-10:00	**Session A.7: Functional Time Series Analysis and Application**	
10:05- 11:20	**Session A.8: Time Series Analysis with Computational Intelligence (part II)**	**Session B.4: Advanced in Time Series Analysis and Forecasting. Short Presentation (part II)**
11:20-12:00	**COFFEE BREAK**	
12:00-12:55	**Session A.9: Forecasting Complex/Big data**	**Session B.5: Econometric Models. Short Presentation**
13:00-14:00	**Session A.P2: Plenary Lecture.** **Prof. Martin Wagner** **University of Klagenfurt, Chief Economic Advisor at the Bank of Slovenia**	
14:00-16:00	**REST BREAK**	
16:00-17:05	**Session A.10: Advanced in Time Series Analysis and Forecasting (part I)**	**Session B.6: Applications in: energy, finance, transportation, networks, meteorology, health, research and environment, etc (part II)**
17:10-18:25	**Session A.11: Econometric Forecasting**	**Session B.7: Time Series Analysis with Computational Intelligence**
18:30-19:50	**Session A.12: Advanced in Time Series Analysis and Forecasting. Short Presentation (part II)**	
21:00	**GALA DINNER** **Hotel Lopesan Baobab 5*** **(15 minutes walking from Hotel Lopesan Villa del Conde Resort)**	

Thursday, June 30th, 2022		
9:00	**REGISTRATION DESK** ***(start at 9:00h but it is opened during all the conference)***	
	All **Sessions A: Oral face-to-face sessions**. All **Sessions B:** Oral (will be held on-line by Zoom)	
9:15-10:00	**Session A.13: Statistical methods for Time Series Analysis and Financial Forecasting**	**Session B.8: Forecasting performance evaluation**
10:05- 11:20	**Session A.14: Advanced in Time Series Analysis and Forecasting. Short Presentation (part III)**	**Session B.9: Advances in Neural Networks for Time Series Forecasting**
11:20-12:00	**COFFEE BREAK**	
12:00-12:55	**Session A.15: Advanced in Time Series Analysis and Forecasting (part IV)**	**Session B.10: Time Series Analysis and Forecasting**
13:00-14:00	**Session A.P3: Plenary Lecture.** **Prof. Wesley S. Burr** **Trent University**	

Preface

We are proud to present the set of final accepted contributions for the 8th International conference on Time Series and Forecasting (ITISE 2022) held in Gran Canaria (Spain) during June, 27th-30th, 2022.

The ITISE 2022 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, in the field of time series analysis and forecasting.

The aim of ITISE 2022 is to create a friendly environment that could lead to the establishment or strengthening of scientific collaborations and exchanges among attendees, and therefore, ITISE 2022 solicits high-quality original research papers (including significant work-in-progress) on any aspect related to time series analysis and forecasting, in order to motivate 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.

2. **Advanced methods 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 and global forecast.
   - Techniques for dimension reduction.
   - Multiscaling.
   - Forecasting complex/big data.

4. **Forecasting in real problems**.

   - 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.

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 reviewers). In these proceedings we are presetting the abstract of the contributions to be presented during ITISE-2022 (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 speakers:

Prof. Daniel Peña Sánchez De Rivera, Professor at Universidad Carlos III de Madrid. Department of Statistics. Madrid (Spain).

Daniel Peña was born in Madrid in 1948. He received his PhD in Industrial Engineering from the Polytechnic University of Madrid, Bachelor's Degrees in Sociology and Statistics from the Complutense University of Madrid and Business Administration ITP from Harvard. President of Carlos III University of Madrid in the period 2007-2011 and re-elected in March 2011. He was Director of the Management Committee (1993-2000) and Vice Rector of the Carlos III University of Madrid (1992- 1995), where he is Full Professor of the Department of Statistics. He has also been Full Professor at the Polytechnic University of Madrid, the University of Wisconsin- Madison and the University of Chicago. He is Founding Director of the Quantitative Methods Department of the EOI Business School, of the Statistics Laboratory of the ETSII-UPM (the Higher Technical School for Industrial Engineering at the Polytechnic University of Madrid), as well as the Department of Economics and Statistics and Econometrics of the Carlos III University of Madrid. He has been Director of the "Revista Estadística Española" and President of the Spanish Society of Statistics and Operative Research (Sociedad Española de Estadística e Investigación Operativa). In the same field, he stands out as the Founding President of the Statistical Methods Committee of the Spanish Association for Quality and member of the State High Council of Statistics, Vice President of the Inter-American Institute of Statistics and President of European Courses in Advanced Statistics. He has published thirteen books and more than 190 research articles on Statistics, Quality and its Applications. He is Associate Editor of several international journals and has received national and international awards for research. In 2006 he received the Youden Award for the best article published in "Technometrics". He is an honorary member of prestigious international associations such as the Institute of Mathematical Statistics and the American Statistical Association.

Title of the presentation: New approaches in Dynamic Factor Models. This presentation will describe three advances that are useful for building

Dynamic Factor Models for high dimensional time series. The first one is building factor models with cluster structure. Often we have groups specific factors in addition to main global factors, and a methodology to build these models will be reviewed. The second one is that with highly dissagregated data we have to deal with data with tensor structure and recent methods to incorporate these restrictions into a Dynamic factor model will be studied. Third, we often have outliers in the data and robust procedures to carry out the estimation of a dynamic factor model with lags will be reviewed.

Prof. Martin Wagner, Professor of Economics at the University of Klagenfurt, Chief Economic Advisor at the Bank of Slovenia and Fellow of the Macroeconomics and Economic Policy group at the Institute for Advanced Studies, Vienna.

Martin Wagner currently is Professor of Economics at the University of Klagenfurt, Chief Economic Advisor at the Bank of Slovenia and Fellow of the Macroeconomics and Economic Policy group at the Institute for Advanced Studies, Vienna. From October 2017 until end of 2018 he was Chief Economist of the Bank of Slovenia, being on leave from Technical University Dortmund, where we has been Professor of Econometrics and Statistics in the Faculty of Statistics of the Technical University Dortmund from 2012 until 2019. He was educated in Vienna, at the Technical University and the Institute for Advanced Studies, obtaining Diplomas in Mathematics (1995) and Economics (1998), as well as his Doctorate (2000). He obtained his Habilitation in Economics in 2007 at the University of Bern. Martin Wagner has worked at the Technical University of Vienna, the Institute for Advanced Studies in Vienna, the University of Bern and has been Professor of Econometrics and Empirical Economics at the University of Graz before his arrival in Dortmund. Visiting positions have brought him to Princeton University and the European University Institute in Florence. His work has been published, amongst other outlets, in Journal of Econometrics, Econometric Theory, Journal of Applied Econometrics, Econometric Reviews, Econometrics, Oxford Bulletin of Economics and Statistics, Journal of Empirical Finance, Economics of Transition and Ecological Economics.

Title of the presentation: HAC Robust Estimation and Inference

for Long-Run Equilibrium Relationships Rafael Kawka, Faculty of Statistics, Technical University Dortmund, Dortmund, Germany Matthias Vetter, Department of Mathematics, Kiel University, Kiel, Germany Martin Wagner, Department of Economics, University of Klagenfurt, Klagenfurt, Austria, Bank of Slovenia, Ljubljana, Slovenia and Institute for Advanced Studies, Vienna, Austria Abstract: We extend cointegration analysis to a situation where the first differences of the analyzed processes are so-called locally stationary processes (see, e.g., Dahlhaus, 1997) rather than stationary processes. This allows us to model long-run relationships between time series whilst allowing for more or less turbulent or persistent periods in the analysis. As is common in the cointegrating regression literature, we allow for regressor endogeneity and error serial correlation, now both time-varying because of our locally-stationary setup. The paper starts by developing the required functional central limit results for this setting which then allow showing that: First, the OLS estimator is consistent but its limiting distribution is contaminated by second-order bias terms, which differ, of course, from the bias terms arising in the standard context. Second, a localized version of the fully modified OLS estimator, originally considered in Phillips and Hansen (1990) for a standard cointegration setting, leads to a zero mean Gaussian mixture limiting distribution. An important difference to the standard cointegration setting is that fully modified based inference requires something like a HAC-type correction. The theoretical analysis is complemented by a simulation study as well as an empirical illustration with the forward rate unbiasedness hypothesis.

Prof. Wesley S. Burr, Assistant Professor, Statistics. Department of Mathematics, Trent University.

Wesley S. Burr is a statistician and professor at Trent University in Ontario, Canada. He completed his PhD under the supervision of David J. Thomson at Queen's University in Kingston, Ontario in 2012, and then joined Health Canada's Population Studies Division as a Canadian Governmental Laboratories Visiting Fellow. He began a tenure-track position at Trent University in 2016 and was granted tenure and promotion to Associate Professor in 2022. Wesley's work on time series is primarily focused on applications and theory relating to frequency-domain statistics of time series,

and he is interested in problems of imputation, frequency stationarity, and statistics for estimation and detection of signals in time series, as well as applications in environmental epidemiology, space physics, and spatiotemporal modeling of biofluids.

This new edition of ITISE was organized at the Universidad de Granada. 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, Springer, Associate Editor, for their interest in the future editing a book series of Springer from the best papers of ITISE 2022.

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.

June, 2022

ITISE Editors and Chairs
Olga Valenzuela
Fernando Rojas
Hector Pomares
Luis Javier Herrera
Ignacio Rojas

ITISE 2022 PROGRAM

Monday, June 27th 2022

(18:30-20:00) Registration Desk

(18:30-20:00) Upload the presentations to the room's computer (in case you haven't sent them by email).

Tuesday, June 28th 2022

(9:15-10:15) Session A.1: Time Series Analysis with Computational Intelligence (part I)

Recurrent Neural Networks for Forecasting Time Series with Multiple Seasonality: A Comparative Study 1
Grzegorz Dudek, Slawek Smyl and Paweł Pełka

ODIN TS: a tool for the black-box evaluation of time series analytics 1
Niccolò Zangrando, Rocio Nahime Torres, Federico Milani and Piero Fraternali

A combination of Fuzzy techniques and Chow test to detect structural breaks in time series 1
Phuong Truong and Vilem Novak

Early detection of flash-floods using Case-Based Reasoning 2
Enrique Fernández, José Ramón Villar, Alberto Navarro and Javier Sedano

(10:20- 11:20) Session A.2: Advanced Econometric Methods

HAC Robust Estimation and Inference for Long-Run Equilibrium Relationships 2
Martin Wagner, Rafael Kawka and Mathias Vetter

A semi-Markov approach to financial modelling during the COVID19 pandemic 2
Riccardo De Blasis

Asymmetric New Keynesian Phillips Curve for Mexico, 2005Q1-2021Q4..... 2
Eduardo Loría and Raúl Tirado

Analytic Solution to ISO Problem with Bounded Production Capacities 3
Miroslav Pištěk, Elisabetta Allevi, Martin Branda and Rossana Riccardi

(10:20- 11:20) Session B.1: Applications in: Energy, Finance, Transportation, Networks, Meteorology, Health, Research and Environment, etc (part I)

Forecasting Stock Market Volatility using Hybrid of Fuzzy Inference Rules with Descent Method with Haar Wavelet Function: Saudi Arabia 3
Abdullah Alenezy, Mohd Tahir Ismail, Sadam Alwadi, Jamil Jaber and Nawaf Hamadneh

Deep Learning Approaches for Daily Climate Forecasting, Case Study: Malang Indonesia 3
Alif Aditya Wicaksono, Wiwik Anggraeni and Mauridhi Hery Purnomo

A joint probability density function of wind speed and direction for wind power plant analysis..... 3
Guglielmo D'Amico and Filippo Petroni

Forecast-based Financing for Food Security Interventions in Africa: Case Studies in Djibouti, Ethiopia, Lesotho, Madagascar and Niger 4
Asher Siebert, Bohar Singh, M. Azhar Ehsan, Souha Ouni, Yohana Tekeste, Remi Cousin, Igor Khomyakov, Aaron Kaplan and Daniel Osgood

(12:00-12:55) Session A.3: Time Series Analysis of Cryptocurrency (High-Frequency)

Indexed Semi-Markov Model to forecast volatility in cryptocurrency market 4
Filippo Petroni, Guglielmo D'Amico and Flavio Prattico

Cryptocurrency hourly volatility and information measurement of predictability..... 4
Filippo Petroni and Flavio Prattico

Optimal Sampling Frequency of High-Frequency Cryptoasset Time Series - Application to Determining Structural Breaks 5
Damian Zięba

(12:00-12:55) Session B.2: Nonstationarity

Evaluating the long-term impacts of economic or policy shocks among necessity and opportunity entrepreneurs..... 5
Nicola Rubino, Emilio Congregado and David Troncoso

Modelling the Number of Intra-day Stock Transactions using a Novel Time Series Model 5
Yuvraj Sunecher

Costationary Whitenoise Processes and Local Stationarity Testing 5
Alessandro Cardinali

An Application of Neural Networks to Predict COVID-19 Cases in Italy 5
Lorena Saliaj and Eugenia Nissi

(13:00-14:00) Opening Ceremony. Plenary Talk: Prof. Daniel Peña Sánchez De Rivera

Professor at University Carlos III, Department of Statistics, Madrid (Spain).
Title of the presentation: New approaches in Dynamic Factor Models

(16:00-17:10) Session A.4: Applications in: Energy, Finance, Transportation, Networks, Meteorology, Health, Research and Environment, etc

Dynamical models for projecting COVID-19 hospital demand: could a more traditional time series approach do better? 6
Richard Wood

Investigate FIBA World Cup 2019: Evidence Using Advanced Statistical Analysis And Quantitative Tools 6
Christos Katris

Forecasting of COVID-19 Hospital Occupancy Using Differential Flatness ... 7
Benjamin Franz, Alexander Wasserburger, Christoph Hametner and Stefan Jakubek

Inland Areas, Protected Natural Areas and Sustainable Development 7
Antonio Bertini, Immacolata Caruso and Tiziana Vitolo

(16:00-17:00) Session B.3: Advances in Time Series Analysis and Forecasting. Short Presentation (part I)

The Role of labour market institutions as a driver of Self-employment in OECD countries: A Bayesian approach 7
Emilio Congregado, Ana María Rodríguez-Santiago and Concepción Román

Maintenance innovation and big data 8
Javier Gejo-García, Salvador Pérez-García, Sergio Gallego-García and Manuel García-García

Coarse Grain Spectral Analysis for the low-amplitude signature of multiperiodic stellar pulsators 8
Sebastià Barceló Forteza, Javier Pascual-Granado, Juan Carlos Suárez, Antonio García Hernández and Mariel Lares Martiz

A look into the ICT capital adoption 8
Pablo Casas and Antonio Golpe

Causes and consequences of the ICT adoption 8
Pablo Casas and Antonio Golpe

Convergence clubs in the ICT adoption 8
Pablo Casas and Antonio Golpe

(17:05-18:30) Session A.5: Energy forecasting

Stochastic model for microgrid management 8
Salvatore Vergine, Cesar Alvarez Arroyo, Guglielmo D'Amico, Juan Manuel Escan Gonzalez and Lazaro Alvarado Barrios

Understanding the Interrelationship between Energy-Use Behavior and Socio-Demographic Profile 9
Hevar Palani, Juan G. Acosta, Aslihan Karatas and Sybil Derrible

Forecasting Natural Gas Prices with Spatio-Temporal Copulas 9
Sven Pappert and Antonia Arsova

Power grid load forecasting using ridge regression including weather forecasts, reanalyses, terrestrial and satellite weather data 10
Franko Pandžić, Ivan Sudić, Amalija Božiček, Matko Mesar, Bojan Franc, Marija Žmire and Ivan Šturlić

Least–Squares Wavelet Analysis of PS–InSAR and Precipitation Time Series 10
Ebrahim Ghaderpour, Claudia Masciulli, Marta Zocchi, Roberta Marini, Giandomenico Mastrantoni, Francesca Reame, Gianmarco Pantozzi, Niccolo Belcecchi, Gabriele Scarascia Mugnozza and Paolo Mazzanti

Limitation of deep learning algorithm for prediction power consumption 10
Majdi Frikha, Khaled Taouil, Ahmed Fakhfakh and Faouzi Derbel

(18:20-19:30) Session A.6: POSTER SESSION

Evaluation and ranking of urban sustainability based on sustainability assessment by fuzzy evaluation model 11
Maryam Robati and Fatemeh Rezaei

Macroeconomic Forecasting Evaluation of MIDAS Models 11
Nicolas Bonino-Gayoso and Alfredo Garcia-Hiernaux

Effects of Transportation, Commerce, and Spillover on Airbnb 11
Ruggero Sainaghi and Jorge Chica-Olmo

Detection and quantification of wave trends in Southern Italy 12
Francesco Aristodemo, Tommaso Caloiero, Andrea Lira Loarca and Giovanni Besio

Study of the interaction between cereal yield and hydroclimatic factors through Dynamic Factor Analysis 12
Tom Vanwalleghem and María del Pilar Jiménez Donaire

Estimation of cloud fraction by longwave radiation using machine learning .. 12
Tina Andriantsalama, Chao Tang, Remy Ineza Mugenga, Thierry Portafaix, Mathieu Delsaut, Patrick Jeanty, Alexandre Graillet, Girish Kumar Beeharry, Roddy Lollchund, Tyagaraja S. Modelly Cunden and Béatrice Morel

Partitioning of Net Ecosystem Exchange using Dynamic Mode Decomposition and Time Delay Embedding 13
Maha Shadaydeh, Joachim Denzler and Mirco Migliavacca

Singular spectrum analysis and Fisher-Shannon analysis of MODIS Terra satellite time series of Italian urban forests 13
Luciano Telesca, Nicodemo Abate, Farid Faridani, Carmen Fattore, Michele Lovallo and Rosa Lasaponara

Forecasting South African exchange rate using nonlinear models 14
Diteboho Xaba, Amogelang Senosi, Katleho Makatjane and Ishmael Rapoo

Forecasting Costa Rican Solar Energy Production 14
Randall Rojas, Abel Mendez-Porras and Efren Jimenez-Delgado

Inventory Forecasting and Average Fill Rates: Theoretical Insights and Computational Procedures 14
Joanna Bruzda, Jakub Wojtasik and Babak Abbasi

Applying high dimensional time series forecasting for traffic flow prediction . 15
Siham Alaoui and Badih Ghattas

Oil prices and selected pollutant emissions: A multifractal characterization . 15
Milena Kojic, Petar Mitic and Stephan Schlueter

A Comparison Study of Deep Learning Models Combined with Multistep Time Series Forecasting Strategies 16
Zuokun Ouyang, Philippe Ravier and Meryem Jabloun

Sensor-Based Anomaly Detection for Marine Engines with LSTM Auto-Encoders 17
Valerian Cimniak, Martin Brutsche, Michael Grill, André Kulzer and Michael Bargende

Automatization of the features selection process to improve temperature-based solar radiation models in Southern Spain 17
Juan Antonio Bellido-Jiménez, Javier Estévez Gualda and Amanda Penélope García-Marín

Autoencoders for Anomaly Detection in an Industrial Multivariate Time Series Dataset 18
Theodoros Tziolas, Konstantinos Papageorgiou, Theodosios Theodosiou, Elpiniki Papageorgiou, Theofilos Mastos and Angelos Papadopoulos

Wednesday, June 29th, 2022

(9:00-10:00) Session A.7: Functional Time Series Analysis and Application

Long Short-Term Memory Networks for the prediction of Fuel Cell Voltage and Efficiency 18
Martin Angerbauer, Michael Grill and André Casal Kulzer

Using regional climate reanalysis for astronomical site-testing 18
Sebastian Hidalgo, Casiana Muñoz Tuñón, Julio Castro Almazán and Antonia Varela Pérez

T-DPnet: Transformer-based deep Probabilistic network for load forecasting 18
Omar Bouhamed, Manar Amayri and Nizar Bouguila

On the prospective use of deep learning systems for earthquake forecasting over Schumann Resonances Signals 19
Carlos Cano, Ruxandra Stoean, Nuria Novas Castellano, Manuel Fernandez Ros, Gonzalo Joya Caparrós and Jose Antonio Gázquez Parra

(10:05-11:20) Session A.8: Time Series Analysis with Computational Intelligence (part II)

Reconstructed Phase Spaces and LSTM Neural Network Ensemble Predictions 19
Sebastian Raubitzek and Thomas Neubauer

An Open-Source and Reproducible Implementation of LSTM and GRU Networks for Time Series Forecasting 19
Gissel Velarde, Pedro Brañez, Alejandro Bueno, Rodrigo Heredia and Mateo Lopez-Ledezma

Forecasting nonlinear causal relationships between crude oil prices and exchange rates using machine learning methods 20
Witold Orzeszko

APUMPEDI: Approximating Pan Matrix Profiles of Time Series Under Unnormalized Euclidean Distances by Interpolation 20
Jing Zhang and Daniel Nikovski

Macroeconomic Predictions using Payments Data and Machine Learning ... 20
Ajit Desai and James Chapman

(10:05-11:20) Session B.4: Advances in Time Series Analysis and Forecasting. Short Presentation (part II)

Outliers impact on parameter estimation of Gaussian and non-Gaussian state space models: a simulation study 21
F. Catarina Pereira, A. Manuela Gonçalves and Marco Costa

Linear Programming-Based Bi-Objective Optimization for Short Univariate Time Series Forecasting 21
Santhosh Kumar Selvam, Chandrasekharan Rajendran and Ganesh Sankaralingam

Exploring the performance of Test Statistics of Time Series Outlier Detection: Evidence from Robust Estimation Methods 22
Tanzeela Yaqoob and Arfa Maqsood

Combining ARIMA-SVR and Fuzzy Time Series Model based on Hedge Algebra to Forecast Bitcoin 22
Dinh Thuan Nguyen and Tung Hoang

Big data oriented to innovation through quality techniques 22
Salvador Pérez-García and Manuel García-García

Time Series Clustering of High Gamma Dose Rate Incidents 22
Mohammed Al Saleh, Béatrice Finance, Yehia Taher, Ali Jaber and Roger Luff

A Hybrid Model of VAR-DCC-GARCH and Wavelet Analysis for Financial Markets Volatility 23
Maryam Nafisi Moghadam and Shahram Fattahi

Multiplicative Nonstationary Volatility Models with Exogenous Information 23
Cristina Amado and Timo Teräsvirta

(12:00- 12:55) Session A.9: Forecasting Complex/Big data

Unbiased estimation of autoregressive models based on the SIMEX algorithm 23
Wilfredo Palma, Susana Eyheramendy, Felipe Elorrieta and Esteban Barrientos

Deep Learning Applied to Wind Power Forecasting: a Spatio-Temporal Approach 23
Rubén del Campo, Eloy Anguiano, Álvaro Romero and José R. Dorronsoro

Dynamical Component Analysis: Updated and improved algorithm, applications, and limitations 24
Monika Warmuth, Philipp Romberger, Knut Hüper and Christian Uhl

Traffic flows time series clustering in flooding risk areas using origin-destination mobile phone data 24
Maurizio Carpita, Giovanni De Luca, Rodolfo Metulini and Paola Zuccolotto

(12:00- 12:55) Session B.5: Econometric Models. Short Presentation

Modelling and predicting the dynamics of confirmed Covid-19 cases based on climate data 25
Yuzhi Cai, Fangzhou Huang and Jiao Song

Dynamic Asymmetric Causality Tests with an Application 25
Abdulnasser Hatemi-J

The implications of the ICT adoption for wages and employment 26
Pablo Casas and Concepción Román

Time series analysis of topological indicators of the network of connections in the insurance sector during the greatest financial crises of the 21st century. 26
Anna Denkowska and Stanisław Wanat

Renewable energy transition in Europe: a multi-national analysis. 26
Andrea Savio and Mariangela Guidolin

Utilising Empirical Mode Decomposition, CNN-LSTM Network with Bayesian Optimisation for Accurate Short-term Residential Load Forecasting 26
Ashkan Lotfipoor, Sandhya Patidar and David P. Jenkins

Demand forecasting under lost-sales stock policies with Tobit Exponential Smoothing 27
Diego J. Pedregal, Juan R. Trapero and Enrique Holgado

(13:00-14:00) Plenary Talk:

Prof. Martin Wagner

University of Klagenfurt, Chief Economic Advisor at the Bank of Slovenia *Title of the presentation: HAC Robust Estimation and Inference for Long-Run Equilibrium Relationships*

(16:00-17:05) Session A.10: Advances in Time Series Analysis and Forecasting (part I)

Online estimation methods for irregular autoregressive models 28
Felipe Elorrieta, Lucas Osses, Matias Caceres, Susana Eyheramendy, Wilfredo Palma and Esteban Barrientos

Noise inference for ergodic Lévy driven SDE 28
Hiroki Masuda, Lorenzo Mercuri and Yuma Uehara

CARMA(p, q) Hawkes processes 29
Lorenzo Mercuri, Andrea Perchiazzo and Edit Rroji

Synthetic subject generation with coupled coherent time series data 29
Xabat Larrea, Mikel Hernandez, Gorka Epelde, Andoni Beristain, Cristina Molina, Ane Alberdi, Debbie Rankin, Panagiotis Bamidis and Evdokimos Konstantinidis

(16:00-17:05) Session B.6: Applications in: energy, finance, transportation, networks, meteorology, health, research and environment, etc (part II)

Renewable Energy Installations and Policy Assessment: Evidence from cointegration and non-causality in Greece (1987-2020) 29
Georgios Evangelidis, Konstantinos Konstantakis and Panayotis G. Michaelides

Forecasting Temperature of Satellites in Low Earth Orbit 30
Kiranmai Madisa, Sharmistha Dey and Paulo de Souza

Hybrid K-Mean Clustering and Markov Chain for Network Accessibility and Retainability Prediction 31
Amel Salem Omer, Tesfaye Addisie Yemer and Dereje Hailemariam Woldegebreal

Air pollution modeling for southern Poland through aggregated land-use regression 31
Katarzyna Kaczmarek-Majer, Olgierd Hryniewicz, Julia Soloch, Krzysztof Skotak and Anna Degórska

(17:10-18:25) Session A.11: Econometric Forecasting

Value-at-Risk Forecasting with Hierarchical Outer Power Archimedean Copulas 31
Jan Gorecki

A Collection of Wisdom in Predicting Sector Returns 32
Hsiu-Lang Chen and Jolana Stejskalová

Relationship between stationarity and dynamic convergence of time series .. 32
Gerardo Covarrubias and Xuedong Liu

The Vector Error Correction Index Model: Representation and Statistical Inference 32
Gianluca Cubadda and Marco Mazzali

Inflation Expectations, Inflation Target Credibility and the COVID-19 Pandemic: New Evidence from Germany 32
Dieter Nautz

(17:10-18:25) Session B.7: Time Series Analysis with Computational Intelligence

Online Classification of High Gamma Dose Rate Incidents 33
Mohammed Al Saleh, Béatrice Finance, Yehia Taher, Ali Jaber and Roger Luff

Multichoice Entropy Clustering for Time Series and Seasonality 33
Miguel Ángel Ruiz Reina

Do News Sentiments Predict Stock Price Direction? 33
Fennee Chong and Bharanidharan Shanmugam

Combination of post-processing methods to improve high resolution solar irradiance forecasts in French Guiana 33
Rafael Alvarenga, Hubert Herbaux and Laurent Linguet

Time Series sampling 34
Florian Combes, Ricardo Fraiman and Badih Ghattas

(18:30-19:50) Session A.12: Advances in Time Series Analysis and Forecasting. Short Presentation (part II)

Measuring the impact of monetary policy on the stock market via an identified multivariate GARCH model 34
H. Herwartz, J. Roestel and F. Xu

Point and probabilistic forecast reconciliation for general linearly constrained multiple time series 34
Daniele Girolimetto and Tommaso Di Fonzo

The state of Africa's air transport market amid COVID-19, and forecasts for recovery 35
Tassew Tolcha

Rethinking Tourism and Hospitality Research 36
Mireya Morgana-Orellana, Patricia Picazo-Peral and Sergio Moreno-Gil

Using Forecasting Methods on Crime Data: The SKALA Approach of the State Office of Criminal Investigation North Rhine-Westphalia, Germany. . . . 36
Kai Seidensticker and Katharina Schwarz

Machine learning and satellite data processing for spatio-temporal forecasting - application to solar energy in Guyana 38
Sara Zermani, Jessica Bechet and Laurent Linguet

The (a)symmetric aspects of business-consumer confidence on outward-inward tourism for Russia 38
Magdaline Enow Mbi Tarkang Mary

Exploration of Different Time Series Models for Soccer Athlete Performance Prediction 38
Siarhei Kulakou, Cise Midoglu, Nourhan Ragab, Matthias Boeker and Paal Halvorsen

A Dynamic Combination of Optimized Theta Method with ATA, and Exponential Smoothing Methods; Validating on a Real Business Case and Comparing to Moving Average 39
Yasin Tadayonrad and Alassane Ballé Ndiaye

Thursday, June 30th, 2022

(9:15-10:00) Session A.13: Statistical methods for Time Series Analysis and Financial Forecasting

Cloud Base Height estimation based on Convolutional Neural Network and All Sky Images 40
Emanuele Ogliari, Alfredo Nespoli, Dario Ronzio and Elena Collino

Alone we can do so little; together we can not be detected. Detection of outliers in time series by analyzing cluster transitions. 40
Sergej Korlakov, Gerhard Klassen, Stefan Conrad and Marcus Bravidor

Comparison of different portfolio optimization strategies to minimize the risk 40
Aleš Kresta and Anlan Wang

(9:15-9:45) Session B.8: Forecasting performance evaluation

Recommendations of stockbrokers vs. fuzzy portfolio approach in construction sector 41
Aleksandra Wójcicka-Wójtowicz and Anna Łyczkowska-Hanćkowiak

An approximation to relative automatic time series forecasting measures 42
Miguel Ángel Ruiz Reina

(9:50-11:20) Session A.14: Advances in Time Series Analysis and Forecasting. Short Presentation (part III)

High-Dimensional Temporal Disaggregation and Nowcasting: Examination of Trade-in-Services throughout the Brexit Transition Period 42
Luke Mosley, Alex Gibberd and Kaveh Salehzadeh Nobari

Estimating Growth at Risk with Skewed Stochastic Volatility Models 43
Elias Wolf

Performance Evaluation of Neural Network-Based Short-Term Solar Irradiation Forecasts 44
Simon Liebermann, Jung-Sup Um, YoungSeok Hwang and Stephan Schlüter

Effects of Covid on the performance of different prediction methods: fare revenue forecasting in public transportation in Berlin 45
Nicki Lena Kämpf, Jonas Krembsler, Sandra Spiegelberg, Nicola Winter, Thomas Winter and Robert Knappe

Performance evaluation of a family of GARCH processes based on Value-at-risk forecasts 45
Alex Babiš and Beáta Stehlíková

On the correlation analysis of illiquid stocks 46
Hamdi Raissi

Filling the Gap between Continuous Time AR Processes and Discrete Time AR Sequences. Application to Time Change Issues. 46
Rachid Senoussi and Valérie Girardin

Bridging fundamental and statistical models for short term electricity forecasting for the German market. 47
Souhir Ben Amor, Thomas Möbius and Felix Müsgens

PV fault diagnosis method based on time series electrical signal analysis 47
Carole Lebreton, Fabrice Kbidi, Frédéric Alicalapa, Michel Benne and Cédric Damour

An Example of Time-Series Modelling with Bayesian Symbolic Regression .. 47
Krzysztof Drachal

Expectation-maximization algorithm for autoregressive models with Cauchy innovations 49
Monika Singh Dhull and Arun Kumar

Towards time series feature engineering in automated machine learning for multi-step forecasting 49
Can Wang, Mitra Baratchi, Thomas Thomas Bäck, Holger H. Hoos, Steffen Limmer and Markus Olhofer

(10:05-11:20) Session B.9: Advances in Neural Networks for Time Series Forecasting

Improving the Predictive Power of Historical Consistent Neural Network 49
Rockefeller Rockefeller, Bubacarr Bah, Vukosi Marivate and Hans-Georg Zimmermann

Multi-Country and Multi-Horizon GDP Forecasting Using Temporal Fusion Transformers 50
Juan Laborda, Sonia Ruano and Ignacio Zamanillo

Deep Representation Learning for Cluster Level Time Series Forecasting 50
Tsegamlak Terefe Debella, Bethelhem Seifu Shawel, Maxime Devanne, Jonathan Weber, Dereje Hailemariam Woldegebriel, Sofie Pollin and Germain Forestier

On Generating Synthetic Time Series For Cybersecurity 50
Veronica Sirghei

Time Series Forecasting Using Smoothing Ensemble Empirical Mode Decomposition and Machine Learning Techniques 51
Mona Bahri, Sahand Vahidnia and Mahtab Mohtashamkhani

(12:00- 12:55) Session A.15: Advances in Time Series Analysis and Forecasting (Part. IV)

Heterogeneous agent model of multi-assets in discrete-time 51
Serena Brianzoni, Giovanni Campisi, Graziella Pacelli and Filippo Petroni

An ordinal procedure to detect change points in the dependence structure between non-stationary time series 52
Alexander Schnurr and Svenja Fischer

Modelling a Continuous Time Series from FOU(p) processes 52
Juan Kalemkerian

(12:00- 12:55) Session B.10: Time Series Analysis and Forecasting

Nowcasting US-Inflation with random forests 52
Wegmüller Philipp

Three states of the French business cycle 52
Catherine Doz and Anna Petronevich

Short-Term Electricity Load Forecasting: Modelling the COVID-19 Lockdown in Spain 53
Eduardo Caro Huertas and Jesús Juan Ruiz

(13:00-14:00) Plenary Talk: Prof. Wesley S. Burr

Trent University, Department of Mathematics

Virtual Session

Testing for and measuring serial dependence by neural networks 53
Jinu Lee

Measuring the dynamic cross-correlation and contagion effect between Brent crude and heating oil (US-Diesel) prices pre and during the Covid-19 outbreak 53
Claudio M. C. Inacio Jr and Sergio A. David

Forecasting Macroeconomic Tail Risk with Big Data Quantile Regressions .. 53
Jan Prüser and Florian Huber

Improve the forecasting accuracy using GJR-GARCH and MODWT (C6) .. 54
Tariq Alshammari, Mohd Tahir Ismail, Sadam Alwadi, Mohammad H. Saleh and Jamil Jaber

Using Multi-Layer Perceptron (MLP) to Detect Outliers in Saudi Arabia Stock market 54
Khudhayr Rashedi, Mohd Tahir Ismail, Sadam Alwadi, Jamil Jaber and Nawaf Hamadneh

The bootstrap for testing the equality of two multivariate stochastic processes with an application to financial markets 54
Ángel López-Oriona and José A. Vilar

Bootstrap-based nonlinear shrinkage of sample eigenvalues 55
Eran Raviv

Forewarning pest attacks in crops using beta regression 55
Bishal Gurung, Kn Singh, Achal Lama and Biwash Gurung

The WEF's Global Competitiveness Reports: A case of reliable nonsense? .. 55
Harald Bergsteiner

Probabilistic Forecasting for Oil Producing Wells using Seq2seq Augmented Model 56
Hadeel Afifi, Mohamed Elmahdy, Motaz El Saban and Mervat Abu-Elkheir

Geothermal time series analysis of the Deception volcano 2011-2022 (Antarctica) 56
Belen Rosado, Luis Miguel Peci, Amós De Gil, Paola Barba and Manel Berrocoso

An analysis of liquidity through the self-similarity surfaces 56
Sergio Bianchi, Augusto Pianese and Massimiliano Frezza

Analysis of underwater temperature series of Deception and Livingston Islands 2012-2022 (Antarctica) 57
Olga Luengo, Belén Rosado, Amós De Gil, Paola Barba and Manuel Berrocoso

Analysis of tidal series of Deception and Livingston Islands (Antarctica) 57
Olga Luengo, Belén Rosado, Amós De Gil, Paola Barba and Manuel Berrocoso

Influence of the Eruption of La Palma on the Evolution of Tropospheric Activity in the Canary Archipelago 57
Paola Barba, Belén Rosado, Javier Ramírez-Zelaya, Adrián Villanueva and Manuel Berrocoso

Economic and Legal Nature of Blockchain 58
Farouq Alazzam and Iryna Diachuk

Impact of capitalization on labor productivity in European countries 59
Antonio Mihi-Ramírez, Elías Melchor-Ferrer and Ammar Ziad-Alwrekiat

LASSO regression: an application for Brazilian hydrological time series 61
Johann Freitas

Forecasting Net Asset Value (NAV) for Exchange Traded Fund(ETF) using Time Series ARIMA Model 61
Zobia Mushtaq and Muhammad Rafi

Economic uncertainty impact in forecasting Uruguayan macro variables 61
Bibiana Lanzilotta, Gabriela Mordecki and Viviana Umpiérrez

A decoder-based transformer model for multi-step fresh produce price forecasts 62
Fatemeh Jafari, Kumaraswamy and Fakhri Karray

K-means Clustering Assisted Spectrum Utilization Prediction with Deep Learning Models 62
Bethelhem S. Shawel, Frehiwot Bantigegn, Tsegamlak Terefe Debella, Sofie Pollin and Dereje H. Woldegebreal

A Multivariate Approach for Spatiotemporal Mobile Data Traffic Prediction 62
Bethelhem S. Shawel, Endale Mare, Tsegamlak Terefe Debella, Sofie Pollin and Dereje Hailemariam Woldegebreal

Comparative analysis of residential load forecasting with different levels of aggregationComparative analysis of residential load forecasting with different levels of aggregation 63
Ana Apolo Penaloza, Roberto Leborgne and Alexandre Balbinot

Ensemble Forecasting by Energy associated modified by Renyi's entropy and statistical roughness in the learning process 63
Cristian Rodriguez Rivero, Julián Pucheta and Martin Herrera

Feature Importance Study on Forecasting Hourly Retail Time Series 63
Bruno Mendes, Ana Maria Tomé and José Moreira

Evaluating a Recurrent Neural Networks Model For Predicting Readmission to Cardiovascular ICUs Based on Clinical Time Series Data . . . 64
Sobhan Moazemi, Sebastian Kalkhoff, Steven Kessler, Zeynep Boztoprak, Vincent Hettlich, Artur Liebrecht, Roman Bibo, Bastian Dewitz, Artur Lichtenberg, Hug Aubin and Falko Schmid

Recurrent Neural Networks for Forecasting Time Series with Multiple Seasonality: A Comparative Study

Grzegorz Dudek, Slawek Smyl and Paweł Pełka

Abstract: This paper compares recurrent neural networks (RNNs) with different types of gated cells for forecasting time series with multiple seasonality. The cells we compare include classical long short term memory (LSTM), gated recurrent unit (GRU), modified LSTM with dilation, and two new cells we proposed recently, which are equipped with dilation and attention mechanisms. To model the temporal dependencies of different scales, our RNN architecture has multiple dilated recurrent layers stacked with hierarchical dilations. The proposed RNN produces both point forecasts and predictive intervals (PIs) for them. An empirical study concerning short-term electrical load forecasting for 35 European countries confirmed that the new gated cells with dilation and attention performed best.

ODIN TS: a tool for the black-box evaluation of time series analytics

Niccolò Zangrando, Rocio Nahime Torres, Federico Milani and Piero Fraternali

Abstract: Time series datasets and tasks such as anomaly detection and predictive maintenance are gaining importance due to the progresses in Deep Neural Networks and to the diffusion of IoT applications. The analysis of performance for these tasks relies on standard metrics applied to the entire dataset. Such indicators provide a global performance assessment but might not help a deep understanding of the model weaknesses. A complementary diagnostic approach exploits error categorization and ad-hoc visualizations. In this paper we present ODIN, an open source diagnosis framework for time series analysis that lets developers compute performance metrics, disaggregated by different criteria, and visualize diagnosis reports. ODIN is agnostic to the training platform and can be extended with application- and domain-specific meta-annotations and metrics with almost no coding. We show ODIN at work through two time series analytics examples.

A combination of Fuzzy techniques and Chow test to detect structural breaks in time series

Phuong Truong and Vilem Novak

Abstract: A new methodology for the detection of structural breaks in time series is introduced. The proposed method is based on the combination of special techniques of fuzzy modeling, namely Fuzzy Transform (F-transform) and selected methods of Fuzzy Natural Logic (FNL), and the Chow test that is the well-known statistical method for anomaly detection in time series. The concept of detection structural breaks in time series using the Chow test is based on the difference of intercept terms as well as different regression coefficients. The time series in which a structural break occurs is divided into two linear regressions models. The test of hypothesis related to the test of structural break is conducted by testing the null hypothesis. However, it is difficult to determine the changepoint in a given time series. Therefore, we apply fuzzy techniques for this purpose. Under the assumption that a time series can be additively decomposed into trend-cycle, seasonal component, and random disturbances, the fuzzy transform makes it possible to find the arbitrary shapes of the trend-cycle, and detect specific areas, or intervals of monotonous behavior. We can also estimate the slope (average value of the first derivative) of the time series in a given (imprecisely determined) area. FNL provides a mathematical model of semantics of special expressions of natural language (the so-called evaluative linguistic expressions). Using them we are able to evaluate the slope of time series or detect areas of monotonous behavior. Using this information, we are able to detect possible structural breaks. The fuzzy transform used in time series provides zero degree (absolute components) and first degree ones. The coefficients are estimations of average values of the tangents (slopes) of the function over areas characterized by the fuzzy sets. The structural break divides set of component into 2 subsets, that can be approximated by linear models. Using Chow test, we confirm presence of the structural break by testing of the null hypothesis. We will demonstrate effectiveness of the mentioned methodology on several real time series.

Early detection of flash-floods using Case-Based Reasoning

Enrique Fernández, José Ramón Villar, Alberto Navarro and Javier Sedano

Abstract: The sudden raise of the water level in a basin due to an abrupt change in the weather conditions is known as flash-flood. The importance of an early flash-flood detection is given by the reduction in its consequences, either in infrastructure damage or in human losses. Up to now, the studies in the literature focus on the dynamics of the basins, determining how the water levels in a basin would be in a considered scenario. Still, the early and on-line flash-flood detection has not been addressed. To cope with this problem, this research proposes a Case-Based Reasoning that estimates the flooding map for a certain prediction horizon. Provided enough data is available, this CBR tool would perfectly deal with different basins and locations. This research is being designed, developed and tested on two real basins, one from Spain and one from France. Expectancy is that the performance of the CBR tool will perfectly mimic the decision making of the public safety experts.

HAC Robust Estimation and Inference for Long-Run Equilibrium Relationships

Martin Wagner, Rafael Kawka and Mathias Vetter

A semi-Markov approach to financial modelling during the COVID19 pandemic

Riccardo De Blasis

Abstract: The general approach to the study of the financial time series is mostly based on the application of the econometric tools of the time series analysis, also called the macro-to-micro approach. However, in recent years, a new strand of the literature has emerged. This new area deals with those problems looking at the opposite perspective, namely the micro-to-macro approach, which directly models the observable quantities. With the macro-to-micro approach, the observed price is considered to be a noisy representation of an unobserved price. On the contrary, the micro-to-macro approach focuses only on the observable quantities and exploits the point processes (Fodra and Pham, 2015 DOI 10.1080/1350486X.2015.1037963). Among this new area of the literature, the Weighted-Indexed Semi-Markov model (WISM) by D'Amico and Petroni (2012, DOI 10.1016/j.physa.2012.05.040), and its extensions (D'Amico and Petroni, 2018 DOI 10.1016/j.ejor.2017.12.016, 2021 DOI 10.1002/asmb.2622), has proven to be able to reproduce important stylised facts of financial time series such as the first-passage-time distributions and the persistence of volatility. The WISM model is a generalisation of the semi-Markov model. While the latter employs two random variables, the observed price returns and the time between each price change, the WISM includes a third variable that takes into account the history of the time series and their intercurrent time, thus allowing for better reproduction of the observed quantities. The application of the WISM model to financial time series requires the conversion of the continuous returns into a discrete state space. In addition, the inclusion of the index that captures the history of the process requires a further discretisation. At the moment, both conversions are based on the visual inspection of the distribution of the processes, imposing a subjective choice. In this paper, we test the effectiveness of an automatic discretisation using the K-means algorithm. This approach permits to reduce the discretion, thus limiting the choice to the number of states. Moreover, the WISM model has been introduced only recently in the literature and its applications are limited to the stock markets series. Therefore, it is useful to test the validity of the model and specifically, its capability of reproducing the typical stylised facts of financial time series when applied to other financial time series. In particular, considering the latest financial disruptions caused by the COVID19 pandemic, we aim to assess the model capabilities under this period of financial distress, which experienced extremely high volatility.

Asymmetric New Keynesian Phillips Curve for Mexico, 2005Q1-2021Q4.

Eduardo Loría and Raúl Tirado

Abstract: The recovery from the corona crisis has raised inflation around the world above central bank targets. In Mexico, this has caused a restrictive monetary policy. We estimate three Augmented Asymmetric Accelerationist Phillips Curves with Labor Precariousness (2005Q1-2021Q4) with NLARDL and made an in-sample fore-cast to choose the best explaining model. We establish that there are significant short and long-run

asymmetric effects of the unemployment gap and the rate of critical labor conditions over inflation. The predominant short-run effect comes from increasing unemployment, but in the long term, the opposite occurs. We found no econometric evidence that negative unemployment gaps raise inflation.

Analytic Solution to ISO Problem with Bounded Production Capacities

Miroslav Pištěk, Elisabetta Allevi, Martin Branda and Rossana Riccardi

Forecasting Stock Market Volatility using Hybrid of Fuzzy Inference Rules with Descent Method with Haar Wavelet Function: Saudi Arabia

Abdullah Alenezy, Mohd Tahir Ismail, Sadam Alwadi, Jamil Jaber and Nawaf Hamadneh

Abstract: This study aims to model and enhance the forecasting accuracy of Saudi Arabia stock exchange (Tadawul) data patterns using the daily stock price indices data with 2026 observations from October 2011 to December 2019. This study employs a nonlinear spectral model of maximum overlapping discrete wavelet transform (MODWT) with Haar mathematical functions in conjunction with fuzzy inference rules with descent method (FIR.DM). We have selected oil price (Loil) and repo rate (Repo) as input values according to multicollinearity test. The input variables in this study have been collected from Saudi Authority for Statistics and Saudi Central Bank. The output variable is obtained from Tadawul. The performance of the proposed model (MODWT-Haar-FIR.DM) is evaluated in terms of mean error (ME), root mean square error (RMSE), and mean absolute percentage error (MAPE). Also, we have compared the MODWT-Haar-FIR.DM model with traditional models, which are autoregressive integrated moving average (ARIMA) model and FIR.DM model. The obtained results show that the performance of MODWT-FIR.DM is better than that of the traditional models. Therefore, the proposed forecasting model is capable of decomposing in the stock markets.

Deep Learning Approaches for Daily Climate Forecasting, Case Study: Malang Indonesia

Alif Aditya Wicaksono, Wiwik Anggraeni and Mauridhi Hery Purnomo

Abstract: Climate prediction is still a challenge, with many models breaking down after a few days of predictions diverging quickly from reality. Better forecasting of various climate factors is essential for decision-making. Although various methods have previously been applied to climate forecasting, none of the papers used the deep learning approach, especially Gate Recurrent Unit (GRU). The study proposes the GRU to forecast daily temperature, humidity, and rainfall data. The results were compared to other deep learning approaches such as Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Multi-Layer Perceptron (MLP). The experiments were conducted on various scenario data split of train test, namely k-fold cross-validation and split time series 80:20 into 2016-2019 data. Each scenario is executed 16 times. Forecasting performance is measured using three different metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Symmetric Mean Absolute Percentage Error (SMAPE). The experiment results showed that the split time series 80:20 had better performances on GRU models and others. In addition, empirical results showed that GRU was superior to the benchmarking model with an average difference SMAPE of 8,081% in temperature, 4,877% in humidity, and 1,553% in rainfall.

A joint probability density function of wind speed and direction for wind power plant analysis

Guglielmo D'Amico and Filippo Petroni

Abstract: Modelling wind power requires not only a reliable stochastic model of wind speed and accurate conversion of speed into power but also needs to consider auxiliary variables that are dependent on the wind speed process. The most important variable is undoubtedly the wind direction. Consistent research has been produced to jointly model wind speed and direction. The majority of contributions rely on static modelling and the specification of suitable joint distributions. In this paper, we consider a dynamic model based on weighted

indexed semi-Markov chains and the use of copula functions to obtain a multivariate model. The proposed model is applied to wind data recorded at a specific location in Sardinia (Italy). The results show that the model is able to reproduce important stylized facts of wind speed, direction and power series in the univariate and multivariate setting.

Forecast-based Financing for Food Security Interventions in Africa: Case Studies in Djibouti, Ethiopia, Lesotho, Madagascar and Niger

Asher Siebert, Bohar Singh, M. Azhar Ehsan, Souha Ouni, Yohana Tekeste, Remi Cousin, Igor Khomyakov, Aaron Kaplan and Daniel Osgood

Abstract: Many semi-arid regions in Africa are prone to droughts which are often linked with widespread food insecurity and even famine. For years, international aid and development organizations have sought to address these challenges with food aid interventions. However, these efforts have often been a reactive response to an emerging or existing disaster, rather than an anticipatory intervention. In this work, we present a model for forecast-based financing (FbF) to better inform food security interventions, through a partnership between the World Food Programme, the International Research Institute for Climate and Society (IRI) and the meteorological services of five countries in Africa: Djibouti, Ethiopia, Lesotho, Madagascar and Niger. We have developed forecast-based financing tools that utilize a multi-model ensemble approach to forecast seasonal rainfall. It enables users to view historical forecast information and a table of historical years with information regarding observed rainfall, forecasted rainfall, El Nino state and farmer reported bad years and to select thresholds for intervention. Results indicate that the FbF tool seems to capture some of the historical bad years fairly well at a lead time of at least several months, thereby potentially strengthening the ability of the World Food Programme to respond to food insecurity in regions prone to those risks. Further research is underway on forecasting other potential predictands (e.g. rainy day frequency, dry spell frequency, onset of rainy season), and using other model variables as predictors (e.g., 2-meter air temperature, sea surface temperatures, winds, etc.).

Indexed Semi-Markov Model to forecast volatility in cryptocurrency market

Filippo Petroni, Guglielmo D'Amico and Flavio Prattico

Abstract: In this paper, we study the stylized facts of cryptocurrency markets and propose a modelling approach based on weighted indexed semi-Markov chains. First, we describe the model and the methods used for parameters estimation and optimizations. Then, the model is applied to real data. We compare the statistical properties of our proposed and those of real data showing that the model is able to reproduce many stylized facts. Then, we study the ability of the proposed model to forecast the volatility of real data in different historical periods.

Cryptocurrency hourly volatility and information measurement of predictability

Filippo Petroni and Flavio Prattico

Abstract: The main purpose of this work is to investigate the relation between some measures in information theory and the accuracy of volatility forecasting using a model of asset returns. First we highlight the dependence between volatility forecasting and entropy and then we determine the relation between predictability and volatility. The study is conducted using a database of different crypto currency and by computing a hourly mean volatility. To this end we use the standard GARCH approach to model and forecast the hourly volatility. The main result of this paper is the establishment of a relationship between the accuracy of the volatility forecast and the entropy of the time series of cryptocurrency exchange. Since the entropy changes in time, before computing a forecast of the volatility it is recommended to compute the entropy of the time series that furnishes an important indicator on the limit of successive volatility forecast.

Optimal Sampling Frequency of High-Frequency Cryptoasset Time Series - Application to Determining Structural Breaks

Damian Zięba

Evaluating the long-term impacts of economic or policy shocks among necessity and opportunity entrepreneurs

Nicola Rubino, Emilio Congregado and David Troncoso

Modelling the Number of Intra-day Stock Transactions using a Novel Time Series Model

Yuvraj Sunecher

Abstract: This paper focusses on the impact of the Covid-19 on the Stock Exchange of Mauritius (SEM) by modelling the number of intra-day stock transactions of two banks. Hence, a non-stationary bivariate integer-valued autoregressive and moving average of order 1 (BINARMA(1,1)) process with COM-Poisson (CMP) innovations (BINARMA(1,1)CMP) is introduced. The conditional maximum likelihood (CML) approach is used to estimate the model parameters. The novel model is applied on the intra-day trading of two banking stocks.

Costationary Whitenoise Processes and Local Stationarity Testing

Alessandro Cardinali

Abstract: Loosely speaking, a stationary time series is one whose statistical properties remain constant over time. A locally stationary (LS) time series is one whose statistical properties can change slowly over time. As a consequence, such a series can appear stationary when examined close up, but appear non-stationary when examined on a larger scale. This abstract introduces several new results on locally stationary time series that are built using wavelet building blocks. This flavor of local stationarity is an alternative to the locally stationary Fourier processes that were formalized about thirty years ago even if their formulation dates back to earlier work done during the 60' by M. B. Priestley. The presentation will provide a short review of this work and explains some of the reasons for considering the wavelet-based models. We first propose a framework to derive a nonparametric approximation for the unknown cumulative distribution function (CDF) of unobservable locally stationary innovations. We address this task by introducing a new method to produce pseudo innovations whose empirical CDF can be used to approximate the theoretical CDF of unobservable innovations. In order to build these pseudo innovations, we introduce a new class of stochastic processes that we call 'Costationary Whitenoise Processes'. We obtain realizations of these processes as time-varying linear combination of the (possibly) locally stationary data and an auxiliary Gaussian Whitenoise sequence that we can simulate. The time-varying combinations are driven by piecewise constant functions that we estimate as result of a Generalized Method of Moments (GMM) type optimization. Then, we apply the above framework to a non-parametric bootstrap stationarity test for multiscale locally stationary processes. The finite sample properties of this test are assessed through simulations showing that our method successfully controls rejection rates for both Gaussian and non-Gaussian innovations. An empirical analysis based on financial returns shows that, when used on a rolling window, our test can also help to successfully identify well known economic shocks.

An Application of Neural Networks to Predict COVID-19 Cases in Italy

Lorena Saliaj and Eugenia Nissi

Abstract: COVID-19 pandemic has become the greatest worldwide threat, as it has spread rapidly among individuals in most countries around the world. This study concerns the problem of weekly prediction of new Covid-19 cases in Italy, aiming to find the best predictive model for daily infection number in countries with a large number of confirmed cases. We compare the forecasting performance of linear and nonlinear forecasting

models using weekly COVID-19 data for the period between 24 February 2020 until 16 January 2022. We discuss various forecasting approaches, including a Nonlinear Autoregressive Neural Network (NARNN) model, an Autoregressive Integrated Moving Average (ARIMA) model, a TBATS model and Exponential Smoothing on the collected data and compared their accuracy using the data collected from 23 March 2020 to 20 April 2020, choosing the model with the lowest Mean Absolute Percentage Error (MAPE) value. Since the linear models seem not to easily follow the nonlinear patterns of daily confirmed Covid-19 cases, Artificial Neural Network (ANN) have been successfully applied to solve problems of forecasting nonlinear models. The model has been used for weekly prediction of COVID-19 cases for the next 4 weeks without any additional intervention. The prediction model can be applied to other countries struggling with the COVID-19 pandemic and to any possible future pandemics and also help taking better decisions and get more prepared in terms of health care systems.

Dynamical models for projecting COVID-19 hospital demand: could a more traditional time series approach do better?

Richard Wood

Abstract: COVID-19 has placed unprecedented pressure on healthcare systems of even the most developed nations. In many countries, acute health services have become overwhelmed by large numbers of severely ill individuals, necessitating admission triage protocols, the sharing of mechanical ventilators, and patient diverts to purpose-built COVID-19 field hospitals. Projections of future COVID-19 bed demand have proved valuable to shaping government policy and management of local health resources. For governments, 'do nothing' projections facilitate proactive imposition of societal restrictions to avoid substantial COVID-19 incidence. Modelling these implemented 'do something' strategies allows hospitals to improve their forward planning of ward space, personnel rotas, as well as informing the necessity of cancelling planned elective treatments. The Greater Bristol healthcare system serves a one million resident population in South West England. Throughout the pandemic, a dynamical 'SEIR' model has been used to project acute and intensive care COVID-19 bed demand, with results setting ward-level bed assignments at the system's three acute hospitals, as well as influencing the timing and magnitude of localised societal restrictions. The model used is a purpose-built age-structured variant of the established Susceptible-Exposed-Infectious-Recovered (SEIR) structure, which additionally accounts for population flow through extra compartments relating to symptomatic and asymptomatic infection, hospitalisation, and death. Calibration of the model is based upon fitting to the historical time series of bed demand, with Latin hypercube sampling used to generate a set of random samples across the multidimensional parameter space, with selection through maximum likelihood. This talk will provide further detail on the modelling technique employed as well as covering examples of real-life application as the pandemic has evolved – such as predicting the run-off in bed demand following the first wave, estimating the effect of face coverings and social distancing, and projecting the extent to which mass vaccination can mitigate the effects of sequential relaxations to lockdown measures. Finally, application of a basic ARIMA model for short-term forecasting during the Alpha/Kent wave will be described, in advance of a brief discussion regarding the possible merits of COVID-19 forecasting via dynamical modelling versus more traditional time series approaches.

Investigate FIBA World Cup 2019: Evidence Using Advanced Statistical Analysis And Quantitative Tools

Christos Katris

Abstract: The purpose of this study is to investigate quantitatively the tournament of basketball World Cup 2019. At first is identified the performance of the team by considering some inputs and by using Principal Components Analysis (PCA). Then, is identified the contribution of factors (shooting, rebounding, turnover and free-throw factors) to the performance of the teams with the use of mul-tiple regression. Furthermore, is compared in a head-to-head comparison of offense vs. defence in terms of contribution to team performance. Also, are identified other factors which affected the performance again with the use of multivariate regression. Additionally, in order to identify teams which per-formed better than expected, we consider 3 groups of teams and we employ hierarchical k-means clustering. Finally are suggested machine learning models (Random

Forests and Neural Networks) for the prediction of the final position of the teams on the tournament. It follows a procedure to show the superiority of the above methods over the use of power rankings and also for their comparison. This is the first time that such analysis is taking place, teams can benefit from the results, fans can better understand the game and power rankings before tournaments can be enhanced. The analysis can be used in other sport tournaments as well.

Forecasting of COVID-19 Hospital Occupancy Using Differential Flatness

Benjamin Franz, Alexander Wasserburger, Christoph Hametner and Stefan Jakubek

Inland Areas, Protected Natural Areas and Sustainable Development

Antonio Bertini, Immacolata Caruso and Tiziana Vitolo

Abstract: The issue of environmental quality has been at the heart of the debate that has developed at the meetings of the world's highest summits (Earth Summit, Climate Convention, Kyoto Protocol, Agenda 21) and in most of the policies conducted in recent years. The extent of the consequences generated by the impact on the environment and territories, particularly those in the Mediterranean basin, of policies aimed at exploiting rather than protecting environmental resources, has now become apparent in all its graveness. Over the last sixty years, the Italian development model has focused particularly on urban settlements, neglecting the rest of the territory and, in particular, areas far from the traffic flows of goods and people, together with the lack of infrastructure and network connections. At the same time, the environmental consequences of the covid19 pandemic have made it necessary to introduce new measures to reduce the conditions of isolation, even in the face of the population growth in inland areas. In this direction, the great project that goes by the name of "National Strategy for Inland Areas", has launched since 2014, a new development model centred on small settlements to reduce the current conditions of marginalisation. The methodological approach, starting from the state of the art of the existing literature, focuses on the issue of protection and sustainability of marginal areas, both natural and protected, as well as inland areas, in view of the important implications for spatial planning policies, the management of demographic, tourist and other flows, and the revival of local identities. On the basis of the study experiences carried out by the authors, the result is to contribute to the definition of a context analysis that sees the disadvantaged and protected areas projected towards a new development so that, as in the case of more densely urbanised areas, they can be the subject of at least environmental recovery interventions. In particular, starting from the potential expressed by local communities, landscape resources, cultural, immaterial and material heritage, the aim is to provide indications to enhance the unexpressed economic opportunities linked to both agricultural and tourism activities based on sustainability.

The Role of labour market institutions as a driver of Self-employment in OECD countries: A Bayesian approach

Emilio Congregado, Ana María Rodríguez-Santiago and Concepción Román

Abstract: Previous research on the determinants of entrepreneurship provides mixed and non-robust results, suggesting ambiguous relationships between different potential drivers and entrepreneurship. Differences in data and controls, ignoring non-linearities -misleading inference–, multicollinearity and omitted variables bias are potential reasons behind the controversies about the role of some of these entrepreneurship' drivers. This paper explores new strategies for providing new empirical findings and solving, at least partially, this controversy. In particular, we control the model uncertainty -specification problem–, by using a Bayesian model averaging applied to panel data. Following Moral-Benito (2011), the extension of the Bayesian model averaging (BMA) methodology (Fernández et al., 2001b) to a panel data framework, and Crespo Cuaresma (2011) to model the structure for interactions terms, this paper assesses model uncertainty over a large set of determinants, investigate the robustness of determinants between countries and allows for heterogeneous effects depending on the employment protection legislation. Our results provide new evidence about the drivers of entrepreneurship in the OECD countries, and how the relationships are different depending on the strength of the employment protection legislation.

Maintenance innovation and big data

Javier Gejo-García, Salvador Pérez-García, Sergio Gallego-García and Manuel García-García

Abstract: The paper deals with the new capabilities and challenges of data analysis. To do this, first, the development of how these capacities have been developed and in what current state we find ourselves today will be presented. Then, the different statistical approaches providing solutions to them will be presented. Afterwards, it will be presented an assessment of how these capabilities will change the maintenance management and control of industrial assets. Finally, application cases will be presented in which big data is capable of providing innovations and improvements in maintenance, enabling to generate sustainable processes and products in different sectors of activity.

Coarse Grain Spectral Analysis for the low-amplitude signature of multiperiodic stellar pulsators

Sebastià Barceló Forteza, Javier Pascual-Granado, Juan Carlos Suárez, Antonio García Hernández and Mariel Lares Martiz

Abstract: Coarse Grain Spectral Analysis can explain the possible multiscaling nature of the thousands of low-amplitude peaks observed in the power spectra of some pulsating stars. Space-based observations allowed the scientific community to find this kind of structures thanks to their long-duration, high-photometric precision and duty cycle compared to those from ground. Although these time series are far from perfect (outliers, trends, gaps, etc.), we use our own data preprocessing method known as the 2N+1 stage interpolation method to improve the background noise up to a factor 14, avoiding spurious effects. We apply both techniques, the 2N+1 stage method and the CGSA analysis, shedding some light to a real problem on stellar seismology: to find the physical nature of the low-amplitude signature for multiperiodic stellar pulsators.

A look into the ICT capital adoption

Pablo Casas and Antonio Golpe

Causes and consequences of the ICT adoption

Pablo Casas and Antonio Golpe

Convergence clubs in the ICT adoption

Pablo Casas and Antonio Golpe

Stochastic model for microgrid management

Salvatore Vergine, Cesar Alvarez Arroyo, Guglielmo D'Amico, Juan Manuel Esca��o Gonzalez and L��zaro Alvarado Barrios

Abstract: In recent years, we are witnessing an ever-greater demand for energy and a consequent increase in greenhouse gas emissions and the depletion of fossil fuels. In this scenario, microgrids play a fundamental role because they supply sustainable energy and integrate Renewable Energy Sources such as photovoltaic and wind turbines. In this work, the microgrid that we consider is composed by two renewable sources, a photovoltaic plant and a wind turbine, and two controllable sources, a microturbine and a diesel generator. The microgrid is also connected to an energy storage system which charges or discharges power according to its state of charge. The demand is predicted for the next 24 hours through an ARMA model applied on historical data. We apply a homogeneous Makov chain for modeling the photovoltaic power and a non-homogeneous Makov chain for modeling the wind power. We use 10-year historical data of wind speed and photovoltaic power (from 2008/08/01 to 2018/08/01) referred to a location in Sardinia. Our approach has the aim of incorporating the autocorrelation that exists in these phenomena and considering the dependence of the produced power on the hour within a day. In particular, the wind power shows a different transition probability matrix for each hour of

the day and therefore the chain is not homogeneous. The photovoltaic power is studied considering the seasonal trend as base and modeling the residual with the homogeneous Markov chain. We apply these models within the management of a microgrid scheduling problem which has the aim of minimizing the overall production cost. We add stochasticity and uncertainty to renewable sources and we compare the results obtained from the real and the simulated data. We consider two scenarios in which we modify the power and capacity of the battery and we compare the results of 100-day simulations and real data in terms of total cost and unit commitment of the dispatchable units.

Understanding the Interrelationship between Energy-Use Behavior and Socio-Demographic Profile

Hevar Palani, Juan G. Acosta, Aslihan Karatas and Sybil Derrible

Abstract: Building-related sectors use more energy than all other energy consuming sectors. In the United States, the commercial and residential buildings consume about 74% of electricity, 56% of natural gas, and 41.2% of all primary energy (U.S. Department of Energy 2011). Energy consumption behavior of people is one of the key factors affecting buildings' energy consumption. Despite several international agreements to minimize energy consumption, the U.S. Energy Information Administration (2013) projected an increase of 56% in the world's energy consumption by 2040 due to rapid socio-demographic changes and people's energy consumption behavior. Therefore, this study aims to understand the interrelationships between energy-use behavior of people and their socio-demographic characteristics. To achieve this, data is collected from a longitudinal study of two surveys (i.e., before and after being exposed to proposed interventions and incentives) to measure the impact of socio-demographic variables (e.g., age, gender, education level, ethnic background) on energy-use behavior. Then, the interrelationship between energy-use behavior and socio-demographic profile were investigated and analyzed based on the collected data using machine learning techniques. The results indicated that females with income below $75k, age of over 36 years old, and not North African have higher potential of increasing their Motivation level and change their energy-use behavior toward better. Also, females with income of $75k or more, age below 44 years old, and not African American or Hispanic have higher potential to change their energy-use behavior toward better. The findings of this study can help decision-makers (e.g., utility companies, policy makers) propose effective and efficient strategies to reduce CO2 and greenhouse gas emissions in buildings.

Forecasting Natural Gas Prices with Spatio-Temporal Copulas

Sven Pappert and Antonia Arsova

Abstract: This paper aims to model and forecast multivariate time series with a copula-based approach. We model both the temporal and the cross-sectional dependence with copula models. (Marginal behavior is governed by univariate marginal distributions.) This flexible approach allows for non-linear and asymmetric dependence structures, as well as complex marginal structures, such as skewness and heavy-tailedness. We employ different procedures to obtain valuable point forecasts from resulting non-elliptical probabilistic forecasts. Our main focus lies in jointly modeling European short-term future natural gas and commodity prices. The time series contains daily observations from January 2008 until May 2019. Modeling the joint dependence structure amounts to constructing an appropriate K times p dimensional copula model. The time series exhibits non-linear cross-sectional and intertemporal dependence. Volatile behavior of the time series corresponds to heavy-tailed intertemporal dependence structures. Coinciding extreme events correspond to heavy-tailed cross-sectional dependence. Vine-copulas are considered to flexibly model bivariate dependencies. The model is compared with Gaussian and t-copula models. The multivariate probabilistic forecasts are evaluated using the energy score. Univariate probabilistic forecasts are evaluated by the CRPS. Another contribution of this paper is developing a method to obtain sensible point forecasts from non-elliptic probabilistic forecasts. In non-volatile phases, the univariate probabilistic forecasts are approximately elliptical and the expectation value constitutes a sensible point forecast. However, in volatile phases, past values have a highly non-linear influence on the probabilistic forecast. In this case, it is not obvious what constitutes a sensible point forecast. To this end, an artificial neural network is used to predict the best quantile of the probabilistic forecast to use as a point forecast. The ANN-point-forecasts are compared with the expectation value, median and mode of the probabilistic forecast in terms of the RMSE. The advantage of extracting non-trivial point forecasts from

probabilistic forecasts (in comparison to directly point forecasting with an ANN) is that the non-trivial point forecast is still equipped with a parametric confidence interval, directly obtained from the probabilistic forecast.

Power grid load forecasting using ridge regression including weather forecasts, reanalyses, terrestrial and satellite weather data

Franko Pandžić, Ivan Sudić, Amalija Božiček, Matko Mesar, Bojan Franc, Marija Žmire and Ivan Šturlić

Abstract: Power system control requires information regarding network operations available hours, days and even months in advance. Load fore- casting plays a pivotal role in system scheduling and decision making. Having accurate load forecasts can ensure uninterrupted and reliable power supply to customers along with potentially reducing operational costs for power system operators. This paper presents a method of short- term grid load forecasting for the Croatian power system usable for var- ious power system control services, e.g. participation on the electric- ity day-ahead market. Outputs from multiple regression models, each trained for a specific hour of the forecast, were combined to get the complete grid load forecast. Multiple models with the same architecture but with different features were compared to prove that weather data can greatly improve load model performance. Models were trained and tested on data obtained from the Croatian Transmission System Opera- tor and open source ERA5 data provided by Copernicus Climate Change Services.

Least–Squares Wavelet Analysis of PS–InSAR and Precipitation Time Series

Ebrahim Ghaderpour, Claudia Masciulli, Marta Zocchi, Roberta Marini, Giandomenico Mastrantoni, Francesca Reame, Gianmarco Pantozzi, Niccolo Belcecchi, Gabriele Scarascia Mugnozza and Paolo Mazzanti

Abstract: Time series analysis of InSAR (Interferometric Synthetic Aperture Radar) data represents an important tool for monitoring the displacement of the Earth's surface. Multi-temporal InSAR techniques, such as the Persistent Scatterers Interferometry (PS–InSAR), provide the time series of displacement of PS to study their historical behaviors. From a hazard–assessment perspective, the rapid detection of deformation patterns is crucial for identifying the areas that will be affected by damage due to landslides. Understanding the relationship between triggering factors, such as precipitation and the occurrence of mass movements from the interpretation of SAR time series is still a major challenge. Herein, we first describe some of the traditional methods, such as Pearson correlation analysis for investigating whether there is any possible linear dependency between precipitation and ground deformation measurements. Then, we present the time series analysis methods in the Least-Squares WAVElet (LSWAVE) software for analyzing non–stationary time series which may have missing values or different sampling rates. We also show how these methods can be used to investigate possible relations between displacement and precipitation time series that may have different sampling rates without any need for interpolation or resampling.

Limitation of deep learning algorithm for prediction power consumption

Majdi Frikha, Khaled Taouil, Ahmed Fakhfakh and Faouzi Derbel

Abstract: In recent years, it has been observed that electricity consumption has become high due to the use of several domestic applications in the house. On the other hand, there is a trend to use renewable energy in many houses such as solar energy, energy storage system and electric vehicles. For this reason, forecasting household electricity consumption is essential for managing and planning energy use. Power consumption forecasting is a difficult time series prediction problem. Additionally, it is difficult to predict electric power consumption using traditional forecasting techniques because power consumption has erratic trend components, including regular seasonal trends. So several researchers worked on intelligent algorithms like machine learning and deep learning algorithm to solve this problem. In this work, we demonstrate that deep learning algorithms are not always reliable and accurate in predicting power consumption. We assessed the proposed strategy utilizing a household dataset that's freely accessible from the UCI machine learning store, which contains information on electric power consumption between 2006 and 2010. It comprises 2,075,259 values, counting 25,979 missing values representing 1.25% of the whole information, and these lost values are dealt with in the pre-processing step. This dataset contains measurements of electric consumption at a one-minute sampling rate for nearly four years. The mean squared error (MSE) and the root mean squared error (RMSE) are chosen as models evaluation metrics. In our experiments, we split the data into a training and a test set. During the training process, the model

is tuned using a training set; the model function then predicts the output values for invisible data in the test set. This method is preferable for household application, and takes less time in simulation. In this method, 75% of the data is used as a training set, while the remaining 25% is used for testing. We also conducted various experiments on different deep learning models for comparison, such as the LSTM, CNN, GRU and 3_LAYER_LSTM models. When using the aforementioned methods, the prediction models were trained with up to at 25 epochs. We use a python tool in order to test the deep learning model.After the simulation results, we notice that 3_layer_model presents more accurate results than other algorithms. the results are precise and reliable for a prediction power consumption. These results present a prediction of 1 minute as time stamp. But in the real case, it takes a higher time stamp like 15 min and 60 min because they are used in smart meters. For this reason we are testing the best model 3_LAYER_LSTM with different time stamps 5 min , 15 min and 60 min in order to show the reliability and accuracy of this model. After simulation results , it is seen that the performance of the prediction is reduced more than 40% when the timestamp is increased. This greatly affects the reliability of the 3_Layer_LSTM model. We see that when there is a sudden change or a peak in consumption, the model fails to predict the exact value. According to these results, we can say that deep leaning models are not always reliable and accurate for power consumption prediction.

Evaluation and ranking of urban sustainability based on sustainability assessment by fuzzy evaluation model

Maryam Robati and Fatemeh Rezaei

Abstract: Urban sustainability can be considered as a part of sustainable development, which emphasizes the balance between three dimensions of environmental, economic, and social sustainability to improve human well-being and the quality of life. In this study, it was attempted to assess urban sustainability in the most critical, populous, and expanded district of Tehran (district 4) using sustainability assessment by fuzzy evaluation. For this purpose, 52 primary indices which were divided into 8 secondary components were employed. Among them, 18 indices for environmental issues were grouped into three secondary indices as E1, to E3, and 34 indices for economic–social and cultural issues were classified into five secondary indices as H 1 to H5. The results revealed that 56% of the study area was potentially unsustainable and 44% of it had a medium sustainability in terms of environmental aspects. Furthermore, 65% of the study area had a medium human well-being sustainability and 35% of it was potentially unsustainable in terms of economic–social and cultural aspects. Finally, 45% of the study area was found unsuitable in terms of overall urban sustainability. The functional sample proposed in the present study can provide the directors and planners with the required information about the condition of sustainability in the most important district of Tehran.

Macroeconomic Forecasting Evaluation of MIDAS Models

Nicolas Bonino-Gayoso and Alfredo Garcia-Hiernaux

Abstract: In this paper we compare the nowcasting and forecasting performance of some MIDAS models (ADL-MIDAS, TF-MIDAS and U-MIDAS) when predicting the GDP growth of the four main Euro Area economies in the period 2011Q4-2020Q3. To do so, we consider various high frequency indicators, horizons and subperiods, each of the latter with a distinct level of uncertainty. Meta-regressions, with an average error metric as endogenous variable, are estimated to account for potential differences in performance by sample period. The results obtained with the whole sample do not reveal any difference in the predictive performance of the models under comparison. In contrast, the analysis by subperiods shows that ADL-MIDAS and U-MIDAS are more suitable in less uncertain economic contexts, while TF-MIDAS outperforms them in more unpredictable ones. These findings are robust to the forecasting error metric used, RMSFE or MAFE.

Effects of Transportation, Commerce, and Spillover on Airbnb

Ruggero Sainaghi and Jorge Chica-Olmo

Abstract: This study explores the effects of three new groups of location variables on Airbnb listing performance: healthy and sustainable transportation systems (metro, tourist bus and bike-sharing), commerce (food and non-food establishment and shopping centers) and spillover. Besides, the study includes other groups

of control variables: size, contractual terms, rules, host, guest and distance to city center. Three hypotheses are tested, all supposing a positive effect on listing revenue generated by the three additional groups of location variables. The empirical study is carried out in the city of Milan (Italy), a complex destination able to attract (before Covid-19 outbreak) different targets with an important focus on shopping. Four models are proposed and empirical findings support the hypotheses.

Detection and quantification of wave trends in Southern Italy

Francesco Aristodemo, Tommaso Caloiero, Andrea Lira Loarca and Giovanni Besio

Abstract: An analysis of a 41-year long wave time series was conducted in Southern Italy to detect and quantify trends of the main wave parameters adopted for coastal and offshore engineering purposes: significant wave height, Hs, peak period, Tp, and mean wave direction, theta. The high resolution database by DICCA MeteOcean Re-Analysis was used. At yearly and seasonal scales, the trend detection was carried out on the mean values by the Mann-Kendall test, choosing a significance level equal to 90%. The quantification of the increase/decrease in Hs and Tp and of the clockwise/counterclockwise shifts in theta was assessed by the Theil-Sen estimator. For the mean values of Hs and Tp, increasing trends were noticed between Libyan, Levantine and Aegean seas, while these behaviors were detected along different seas for the maximum ones. Clockwise shifts of theta were identified along the northern area of Ionian and Adriatic seas, while counterclockwise one in the Algerian Sea and in the southern Ionian Sea. The obtained results can have an importance for various coastal and offshore operations, in a hot-spot for climate changes such as the Mediterranean basin.

Study of the interaction between cereal yield and hydroclimatic factors through Dynamic Factor Analysis

Tom Vanwalleghem and María del Pilar Jiménez Donaire

Abstract: Drought is one of the major natural hazards affecting agricultural crop production and food security throughout history. Given the importance of droughts, a good understanding and development of adequate models for predicting drought occurrence at short and medium time scales is essential. A large number of indicators have been proposed to monitor and evaluate drought stress, mostly based on the deviation of rainfall from long-term mean, such as SPI or SPEI. However, relatively few studies have evaluated the performance of these studies against plant response or independent yield data. The problem with plant yield is that in addition to drought, also other factor can influence strongly, such as frost during cricitical stages of plant growth, diseases and pests, or even excessive rainfall and flooding damaging crops. but also plant diseases which could be related to humidity or excessive rainfall This motivates a wider, more detailed study of the relation between plant response and hydroclimatic factor. The objective of this study is to use Dynamic Factor Analysis (DFA) as an alternative tool to physically based model of plant growth to explore the relationship between plant greenness (NDVI) in rainfed cereal and climatic variables. This study uses 14 years of NDVI data, that is correlated with minimum and mean temperature, rainfall, evapotranspiration and soil moisture. Analysis was done in R software, using the package “nowcasting”.

Estimation of cloud fraction by longwave radiation using machine learning

Tina Andriantsalama, Chao Tang, Remy Ineza Mugenga, Thierry Portafaix, Mathieu Delsaut, Patrick Jeanty, Alexandre Graillet, Girish Kumar Beeharry, Roddy Lollchund, Tyagaraja S. Modelly Cunden and Béatrice Morel

Abstract: Cloudiness is the key to understand the variability of surface solar radiation, and to improve its predictability, which is particularly important for photovoltaic (PV) electricity production. This study estimates Cloud Fraction (CF) by observations from the BSRN (Baseline Solar Radiation Network) station at the island of La Réunion (55.5°E, 20.9°S) in the Indian Ocean, using a XGBoost machine-learning model. Variables explored are the radiative fluxes measured at the station, namely the downward longwave radiation (LWD) from a CGR4 pyrgeometer, the global and diffuse components of the downward shortwave radiation (SWD) from two CMP22 pyranometers, the direct component of the SWD from a CH1 pyrheliometer, along with some additional meteorological parameters (surface pressure, surface air temperature, and relative humidity) from a WXT530 weather transmitter. CF data are obtained from collocated and coincident observations from a

SkyCamVision all-sky camera within the UV-Indien network. In the first step, a set of quality control (QC) procedures derived from the BSRN recommendations, is used to test the quality of the LWD measurements taken at the station of La Réunion. And because this study is the first one based on the LWD at this station, the LWD data are also compared to CERES (Clouds and Earth's Radiant Energy Systems) satellite-based product. In the second step, a XGBoost model is applied to estimate CF. A 2-year dataset (September 2019-June 2021) with a 5-min time step is used for this study, with 90% of it used for training and 10% for validation and test. The predictors in the model are then the LWD and SWD fluxes and the additional meteorological parameters measured at the BSRN station, and the predictand is the CF measurements from the UV-Indien network. Cross validation is performed for training. Learning curves are used to evaluate the model during the training and the validation. Statistical metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Square Error (RMSE), are also calculated for the final evaluation and model comparison. The first model fitting with the XGBoost default parameter values returns a model with an overfitting issue. To overcome this problem, several tests were performed with different optimization methods such as Principal Component Analysis (PCA), Recursive Feature Elimination (RFE) and feature importance. Each applied method was followed by tuning of hyperparameters. In this study, the results after PCA are not satisfactory as the model performance degrades. In contrast, in the case of RFE, results in three of the tests performed show an increase in the model's performance. Finally, feature importance gives results similar to the best results with RFE. After feature selections and model optimizations, a final score of 92% is achieved, which is comparable to the results from recent studies. We find that surface pressure is not necessary for this application since it makes the XGBoost model more complex. And aerosols may be important to improve the estimation due to its interaction with clouds. Besides the statistical XGBoost model, a physical model known as the Automatic Partial Cloud Amount Detection Algorithm (APCADA) is also applied as a benchmark to estimate CF using LWD, surface air temperature and relative humidity data. This algorithm is based on an enhanced version of the clear-sky index and provides a robust real-time cloud amount. The output of this algorithm can be compared with cloud cover measurement from other high temporal cloud detection systems. A comparison of the CF estimation with these two methods was performed.

Partitioning of Net Ecosystem Exchange using Dynamic Mode Decomposition and Time Delay Embedding

Maha Shadaydeh, Joachim Denzler and Mirco Migliavacca

Abstract: Carbon losses from ecosystems affect climate change. Ecosystem respiration (Reco) represents a major component of the global carbon cycle. Accurate estimation of Reco dynamics is necessary for better understanding of ecosystem-climate interactions and the impact of climate extremes on ecosystems. This paper proposes a new data-driven method for the estimation of the nonlinear dynamics of Reco using the method of dynamic mode decomposition with control input. An extended abstract of this work is attached.

Singular spectrum analysis and Fisher-Shannon analysis of MODIS Terra satellite time series of Italian urban forests

Luciano Telesca, Nicodemo Abate, Farid Faridani, Carmen Fattore, Michele Lovallo and Rosa Lasaponara

Abstract: Urban forests play an important role in ameliorating the quality of urban areas allowing these to be more sustainable, contributing to face the effects of several natural and anthropogenic hazards. In such context, the monitoring of urban forests represents a crucial task aimed at identifying any disturbance that adversely influences vegetation. The use of satellite systems to monitor the status of forests is well consolidated and the availability of increasingly advanced sensors have opened new frontiers like the possibility to identify phenomena and trends of vegetation degradation before such signs are visible. Among the factors of vegetation degradation and plant deseases there are several dangerous pathogenic bacteria that cause loss of biodiversity with also a significant economic impact. Toumeyella parvicornis that is an alien species (present from southern Canada to northern Mexico) has been detected for the first time in Europe at the end of 2014, in Campania (Italy) on Pinus pinea, in the urban area of Naples. Since then, significant outbreaks associated with this parasite strains have been detected in Southern Italy, and today it is present in the Campania and Lazio regions where has devasting effects, so that it is expected that in the next future it can cause the complete loss of large areas

of pinus trees. These critical conditions clearly highlight the need of an early detection of infected plants, and containment strategies based on large and multiscale scale monitoring tools. Several studies suggest the use of remote sensing that are critical to detect the pre-visual stages of the infection, to prevent that infected but asymptomatic trees can contribute to epidemic spread. In this work we study six urban forests located in different sites of Italy (Roma, Torino, Castel Volturno, Castel Porziano, Appia and Milano) by using the time series of five MODIS Terra satellite indices: Evatranspiration (ET), Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Leaf Area Index (LAI) and Land Surface Temperature (LST). The satellite time series span from 2000 to 2021; such period of investigation is long enough to detect anomalous patterns in forest vegetation dynamics that could reveal the onset of possible parasite-linked "stressed" status. The satellite time series were analysed by means of the the Singular Spectrum Analysis (SSA) to filter out the seasonalities. Then, the residual time series were analysed by using the Fisher-Shannon Analysis (FSA). The Fisher-Shannon Analysis is based on the calculation of two quantities, the Fisher Information Measure (FIM) and the Shannon Entropy (SE), both describing the order/organization properties of nonstationary and complex time series. Our results highlight that the urban forest of Castel Voltuno (Southern Italy) has an anomalous behavior in terms of organization/order features of satellite time series. This site is also well known to be strongly affected by the Toumeyella parvicornis, and the anomalous pattern displayed in terms of organization/order could be due to the presence of this parasite that might suggest the FIM and SE as good indicators of parasite-affected status of vegetation.

Forecasting South African exchange rate using nonlinear models

Diteboho Xaba, Amogelang Senosi, Katleho Makatjane and Ishmael Rapoo

Abstract: Autoregressive integrated moving average (ARIMA) is one of the popular linear models in time series forecasting for the past decades. The models were quite successful in numerous applications. However, these models have a shortcoming as they are unable to capture nonlinearity in economic and financial data. Recently researchers forecast with nonlinear models such as Smooth Autoregressive Transition (STAR) and artificial neural networks (ANNs) alternative to the traditional linear models. ARIMA models, STAR and ANNs are often compared with mixed conclusions in terms of the superiority in forecasting abilities. Hence the aim of this paper is to compares the forecasting ability of the STAR model combined with ARIMA and ANN estimates. This new hybrid produced forecasting accuracy of 94% with the power accuracy of 92.5%. The findings of this study are useful for decision makers and financial sector for future use and planning and forecasting USD/ZAR exchange rates. Furthermore, a base for future researchers for conducting studies on emerging markets, have been contributed. These results are also important to risk managers and investors.

Forecasting Costa Rican Solar Energy Production

Randall Rojas, Abel Mendez-Porras and Efren Jimenez-Delgado

Abstract: Solar energy is a renewable electricity generating resource with very low environmental impact. In this paper, we forecast solar energy production in Costa Rica using variety of competing univariate and multivariate models. For example, we explore Vector Autoregressive models using weather and temporal dummy variables, as well other popular methodologies such as Dynamic Harmonic Regressions, Prophet, and Neural Network Autoregressions, among others. We also consider temporal aggregation models such as "thief" (R. Hyndman, 2016) and forecast combination alternatives. The energy production forecasts are then used to anticipate spikes in electricity demand to help us engineer a more optimal (including minimizing costs) deployment of electricity across a suite of energy production plants.

Inventory Forecasting and Average Fill Rates: Theoretical Insights and Computational Procedures

Joanna Bruzda, Jakub Wojtasik and Babak Abbasi

Abstract: We consider inventory forecasting in periodic review order-up-to inventory systems in the presence of commitments concerning the average fill rate. We assume that the performance review period is equal to the replenishment period or includes multiple replenishment periods. We discuss inventory forecasting in this

context within the decision-theoretic framework, which allows us to introduce the corresponding Murphy diagrams, enabling forecast evaluation without focusing on a particular loss function from the family of consistent loss functions. We also introduce data-driven procedures to compute the forecasts of interest in this paper and derive their asymptotic properties.

Applying high dimensional time series forecasting for traffic flow prediction

Siham Alaoui and Badih Ghattas

Abstract: Businesses can use in-store traffic flow prediction in order to maximize revenues by deploying resources in an adaptive manner and also conduct competitive analysis using the predictions. Within the statistics and machine learning literature, traffic flow prediction problems have typically been treated using one-dimensional time series prediction methods where each store may have a dedicated model for prediction. Here we consider the use of high dimensional time series prediction methods that allow the training and use of a single model to predict traffic flow for multiple time series. The key insight we put forth is that the problem of predicting traffic flow across different categories or types of stores can be solved using high dimensional methods that can identify and learn structure between related time series. We apply state-of-the-art (SOTA) high dimensional time series forecasting methods to the problem of predicting traffic flow to multiple stores of brands that fall into a variety of categories e.g. coffee shops, fast food etc. We test the performance of these SOTA methods using several metrics and find that compared to one-dimensional approaches high dimensional methods perform the best for predicting traffic.

Oil prices and selected pollutant emissions: A multifractal characterization

Milena Kojic, Petar Mitic and Stephan Schlueter

Abstract: Introduction. The world needs to reduce emissions as soon as possible to avoid the harmful effects of climate change. Greenhouse gases emitted by fossil fuels significantly contribute to various environmental problems [1]. The amount of greenhouse gases in the atmosphere reached a new high in 2021, with an annual rate higher than the average between 2011 and 2020. [2]. Furthermore, energy is one of the most important drivers of today's technical, economic, and social activities. To put it another way, energy is necessary for a wide range of current social and economic activities, and we can say that, aside from coal and natural gas, oil is one of the most important fossil energy sources. As a result, assessing the relationship between pollutant emissions and oil prices is a prominent professional and scientific area of interest. Oil price increases can cause inflation and stifle economic progress. Even so, rising oil costs reflect increased pollution due to economic, social, and technical progress. Oil prices affect the prices of goods made using petroleum products directly and transportation, industrial, and heating costs indirectly. Because corporations may pass on pro-duction costs to customers, an increase in these costs could affect the prices of a wide range of goods and services. Oil combustion produces a wide range of pollutants, including carbon monoxide, carbon dioxide, sulfur dioxide, nitrogen oxides, and ozone. Objective. The main objective of this paper is to explore the multifractal properties and interdependencies between oil prices and pollutant emissions using multifractal detrended cross-correlation analysis. We focus on two prominent pollutant emissions for this study – carbon monoxide (CO) and sulfur dioxide (SO2). Carbon monoxide is a dangerous air pollutant, but it has a minimal direct impact on the global climate because it is a poor infrared light absorber. Nevertheless, it impacts climate change due to chemical reactions that increase the quantities of methane and ozone in the troposphere. In the end, CO2 is formed when CO is oxidized [3]. Furthermore, the world needs to improve technology for decreasing sulfur dioxide emissions from oil and natural gas combustion. Reduced SO2 emissions will reduce global warming and acid rain, making the earth a healthier place [4]. Methodology. The methodological approach in this study is three-folded. We first conduct preliminary tests on multifractality for each single time series: oil prices and two air pollutants using multifractal detrended fluctuation analysis [5,6]. Then, we apply qualitative tests of the cross-correlations (the cross-correlation statistics and DCCA coefficient) to both time series pairs Oil prices/CO and Oil prices/SO2 [7]. Finally, to examine multifractal behaviors in the power-law cross-correlations in these two time series pairs, we use multifractal detrended cross-correlation analysis [8]. We study metrics such as the generalized Hurst and the scaling exponent and quantify their statistical properties to discover the contributing small and large scales. We also compute the multifractal spectrum, perform a fourth-degree polynomial regression of the multifractal spectra and estimate complexity metrics describing the degree of multifractality

of the processes. Additionally, we investigate the leading causes of the multifractality of each series pair. Data. This research uses daily data for crude oil prices, CO emissions, and SO2 emissions. The period under study consists of nine years from January 1, 2012, to December 1, 2018, spanning N = 2390 days. Crude oil prices (WTI) are taken from [9] and present a Spot Price in Dollars per Barrel. The daily data for carbon monoxide and sulfur dioxide emissions were taken from [10]. Data for CO emissions (from site 060010009 in California, U.S.) is a Daily Max 8-hour CO Concentration, in parts per million (ppm). Data for SO2 emissions (from site 060010011 in California, U.S.) is a Daily Max 1-hour SO2 Concentration, in parts per billion (ppb). Results. The results indicate cross-correlations exist between the oil prices and CO time series pair and oil prices and SO2 emissions time series pair. The fluctuation function is the increasing function of the segments, implying long-range correlations. The significant multifractal characteristics of both pairs are confirmed by computing Hurst exponents, scaling exponents, and multifractal spectrums. The widths of multifractal spectra are significantly larger than zero, proving that the series pairs are multifractal. By comparing the degrees of multifractal spectrums, we find that the cross-correlation of the time series pair oil prices/SO2 emissions time series pair is stronger than that of oil prices/CO emissions. We also perform the analysis on shuffled time series which suggests that multifractality arises from a broad probability density function and different long-term correlations. Conclusions. The results shed new light on the relationship between oil price and pollutant emissions. We found that the multifractal properties are significant and that non-linear cross-correlation between oil prices/CO and Oil prices/SO2 exists. Researchers working with multivariate frequent time series from any scientific field are encouraged to use the proposed algorithm since it increases the detection of long-range correlations and multifractal features of multichannel data. Future research endeavors can apply the suggested algorithm to various real-world complex system time series, such as financial, economic, climatological, and healthcare. As a result, future research should focus on new time series when a deeper understanding of their relationships is required. 1. Bilgili F., Mugaloglu E. & Koçak E. (2020). The Impact of Oil Prices on CO2 Emissions in China: A Wavelet Coherence Approach. In: Shahbaz M., Balsalobre-Lorente D. (eds) Econ-ometrics of Green Energy Handbook. Springer, Cham. https://doi.org/10.1007/978-3-030-46847-7_2 2. World Meteorological Organization. (2021). WMO GREENHOUSE GAS BULLETIN No. 17. The State of Greenhouse Gases in the Atmosphere Based on Global Observations through 2020. World Meteorological Organization and Global Atmosphere Watch. Available at https://library.wmo.int/doc_num.php?explnum_id=10904 3. Dutch S. I. (2010). Encyclopedia of Global Warming. Salem Press, Pasadena, California, USA and Hackensack, New Jersey, USA. ISBN 9781587655647 4. Ward P. L. (2009). Sulfur dioxide initiates global climate change in four ways. Thin solid films, 517(11), 3188-3203. https://doi.org/10.1016/j.tsf.2009.01.005 5. Podobnik B., Grosse I., Horvatić D., Ilic S., Ivanov P.C. & Stanley H.E. (2009). Quantify-ing cross-correlations using local and global detrending approaches, The European Physical Journal B, 71(2) 243. 6. Zebende G.F. (2011). DCCA Cross-correlation coefficient: quantifying the level of cross-correlation. Physica A: Statistical Mechanics and its Applications, 390, 614–618. 7. Kantelhardt J.W., Zschiegner S.A., Koscielny-Bunde E., Havlin S., Bunde A. & Stanley H. (2002). Multifractal detrended fluctuation analysis of nonstationary time series, Physica A, 316 (1), 87–114. 8. Zhou W.X. (2008). Multifractal detrended cross-correlation analysis for two nonstationary signals, Physical Review E, 77, 066211. 9. EIA(a). The U.S. Energy Information Administration: Petroleum & other liquids. Available online: https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm (accessed on February 28 2022). 10. EPA. The United States Environmental Protection Agency: Outdoor Air Quality Data. Available online: https://www.epa.gov/outdoor-air-quality-data/download-daily-data (ac-cessed on March 3 2022).

A Comparison Study of Deep Learning Models Combined with Multistep Time Series Forecasting Strategies

Zuokun Ouyang, Philippe Ravier and Meryem Jabloun

Abstract: This paper compares the deep learning models combined with different multistep forecasting strategies for time series. Context: In recent years, many deep learning models were proposed for multivariate time series forecasting problems and claimed to be promising in different tasks. Nevertheless, their problem formulations can be ambiguous, and they were actually tested in a simple one-step-ahead strategy, which does not apply to real scenarios requiring multistep forecasting. Motivation: To verify their applicability for multistep forecasting, we conducted our experiments by comparing the performance of three well-known deep learning models, i.e., DA-RNN, LSTNet, and TPA-LSTM, under five multistep forecasting strategies, i.e., One-

Step-Ahead, Recursive, Direct, Multi-Input Multi-Output (MIMO), and Multi-Input Several Multi-Output (MISMO) for three forecasting horizons (3, 6, and 12). Methodology: The three models were tested on six datasets with different types of seasonalities, i.e., Complex Seasonal: Electricity; Nonseasonal: NASDAQ 100 Stock and Exchange Rate; Seasonal: Beijing PM2.5 Data, Solar Energy, and Traffic. Results and Discussion: Our results displayed in Tab. 1 reveal that the Recursive strategy performs the worst while the other strategies perform better with tolerable errors compared to the One-Step-Ahead strategy. Less accumulated errors for the other strategies are reported. Besides, the performance of LSTNet falls behind TPA-LSTM and DA-RNN when facing series with zero or complex seasonalities. One explanation could be that the Recurrent-Skip component in LSTNet for capturing seasonal patterns is not suitable in these situations. While TPA-LSTM uses a CNN to capture the temporal patterns, DA-RNN uses an attention mechanism to focus more on relevant time steps. So it is also interesting to note that DA-RNN performs slightly better in many cases, which means the CNN component in TPA-LSTM is weaker than the attention mechanism in DA-RNN in capturing temporal patterns in long sequences. Surprisingly, we noticed that using the MISMO or Direct strategy can potentially damage the model performance for series with complex seasonalities, especially for TPA-LSTM and DA-RNN, which automatically capture seasonal patterns. Both methods show massive error growth when combined with MISMO and Direct. We think this kind of combination can break the continuity of seasonality, thus damaging their intrinsic ability to capture seasonal patterns. Furthermore, we find that a well-designed attention mechanism might help with the input scale variation, noticing that although without the autoregressive component for the changing scale, which LSTNet and TPA-LSTM both include, DA-RNN still gives, in general, the best results. Conclusion: These models constantly suffer from accumulated errors under the Recursive strategy and thus cannot carry out actual multistep forecasting tasks. However, carefully combining them with the MI(S)MO strategy can tackle this problem and thus enables one-step-ahead deep learning models for multistep forecasting.

Sensor-Based Anomaly Detection for Marine Engines with LSTM Auto-Encoders

Valerian Cimniak, Martin Brutsche, Michael Grill, André Kulzer and Michael Bargende

Abstract: In fleet operation of common low-speed 2-stroke marine engines, a variety of sensor signals are used to monitor the functions of the individual subsystems in the powertrain. E.g. cylinder pressures, temperatures in the exhaust gas, oil and cooling circuit as well as on the liner walls are continuously recorded. To verify those signals for several cylinders and engines on board, computer algorithms are helpful because multidimensional capture often exceeds the capability of human perception and staff on board is limited. In addition, however, engines faults have to be avoided at all time because they lead to dangerous situations in open water as well as to expensive delays in the vessels' schedule. The following approach describes a data-driven approach for mapping the behavior of common subsystems from marine engines and anomaly detection of e.g. liner damage based on the corresponding real-world measurement signals and shows various steps from data analysis to preprocessing and modelling. One of the key points will be a transient anomaly detection for time series of multiple sensor signals, where the underlying algorithm is capable to give an estimation of the actual subsystem condition. Therefore, long-term data of different ships will be preprocessed to train an LSTM auto-encoder machine-learning model to detect irregularities. This modeling will help to find engine issues and thereby support e.g. troubleshooting abilities.

Automatization of the features selection process to improve temperature-based solar radiation models in Southern Spain

Juan Antonio Bellido-Jiménez, Javier Estévez Gualda and Amanda Penélope García-Marín

Abstract: Accurate estimations and forecasts of solar radiation values are crucial due to their importance in different fields like agronomy and energy. The geographical density of automated weather stations having pyranometers is low because their installation and maintenance are very expensive. Several methodologies have been tested to estimate and forecast solar radiation using novel temperature-based features using machine learning models such as Multilayer Perceptron and Support Vector Machine with promising results. However, two major aspects are tackled using an inefficient and low uncertainty algorithm (trial and error): hyperparameter tuning and feature selection. In this work, different algorithms to automate the feature selection process, based on mutual information and Pearson correlation, are assessed to estimate daily solar radiation in

Cordoba (South of Spain) using a Multilayer Perceptron model. The performances have been analyzed using the root mean square error and the correlation coefficient

Autoencoders for Anomaly Detection in an Industrial Multivariate Time Series Dataset

Theodoros Tziolas, Konstantinos Papageorgiou, Theodosios Theodosiou, Elpiniki Papageorgiou, Theofilos Mastos and Angelos Papadopoulos

Abstract: In smart manufacturing, the automation of anomaly detection is essential for increasing productivity. Timeseries data from production processes are often complex sequences and their assessment involves many variables. Thus, anomaly detection with deep learning approaches is considered as an efficient and effective methodology. In this work, anomaly detection with deep autoencoders is examined. Three autoencoders are employed to ana-lyze an industrial dataset and their performance is assessed. Autoencoders based on long short-term memory and convolutional neural networks ap-pear to be the most promising.

Long Short-Term Memory Networks for the prediction of Fuel Cell Voltage and Efficiency

Martin Angerbauer, Michael Grill and André Casal Kulzer

Abstract: Fuel cells are once again experiencing an upswing, as they are a possibility for climate-friendly mobility. Nevertheless, the aim is to operate them at the highest efficiency, which is highly dependent on the operating condition. In order to operate fuel cells at the highest possible efficiency continuously, fast-calculating and reliable predictions are essential. One approach to provide these predictions is artificial neural networks (ANN), which are significantly faster compared to phenomenological models. In this work, recurrent neural networks (RNN) are trained with dynamic data of a proton exchange membrane fuel cell (PEMFC). Due to the different time scales of the processes that occur during the operation of the fuel cell, the latest operating state is not sufficient for a precise prediction. Since, for example, the absorption and release of water takes place slowly, earlier states are required in order to consider these processes. Therefore, the choice fell on RNN with long short-term memory cells (LSTM), which are trained using time series of various dynamic operating cycles. Thus, all time scales are regarded in one combined modelt hat offers fast prediction.

Using regional climate reanalysis for astronomical site-testing

Sebastian Hidalgo, Casiana Muñoz Tuñón, Julio Castro Almazán and Antonia Varela Pérez

Abstract: We present a comparative analysis of a long time-series of 17 years of climate parameters with significance in astronomical site test- ing for Teide (OT) and Roque de Los Muchachos (ORM) observatories. The climate parameters have been obtained from a climate reanalyses database with a spatial resolution of 11 km × 11 km, which has been compared with local averaged data using Automatic Weather Stations (AWS) and radiosounding. The results show the reliability of using cli- mate reanalysis databases for site-testing both at the heights of the OT and ORM observatories (boundary layer) and in the free atmosphere. Cli- mate parameters analyzed are the relative humidity, temperature, wind velocity, and wind direction. For the boundary layer, we have obtained full time-series and annual behavior for each parameter. Monthly aver- aged values are given by using a 17 years time period. Both observatories show values in the range 6–70% for relative humidity, 2.6 – 19 °C for tem- perature, and 5 – 12 m/s for wind velocity, the 95% of time; median values are around 25%, 10 °C, and 5 m/s, respectively. No significant difference between OT and ORM in these climate parameters have been found. The study presented here could be also used by any other current or future projects to demonstrate the feasibility of using climate reanalysis for site-testing comparisons.

T-DPnet: Transformer-based deep Probabilistic network for load forecasting

Omar Bouhamed, Manar Amayri and Nizar Bouguila

Abstract: Since electricity cannot be stored in large quantities and must be supplied as needed, it has become critical for electric power companies to forecast demand ahead of time. However, point forecasting appears to

frequently fail these companies because it ignores the big picture (it neglects information on the uncertainty). This paper proposes an encoder-decoder model that takes advantage of the expressiveness of transformer-based encoders to produce probabilistic forecasts, i.e, a distribution over the future predictions. Two real-world datasets are utilized to incorporate the performance of the proposed framework on two different types of data: hourly load data from the power supply company of the city of Johor in Malaysia and hourly load consumption data from one of Grenoble Institute of Technology's buildings. The former represents aggregated data, which makes identifying patterns and trends easier, but the latter was taken from a single building (non-aggregated), which increases the difficulty of forecasts. The model's performance is discussed across multiple time horizons, including 24-hour, 1-week, and 1-month predictions. The framework achieved notable improvements compared to the used baseline, Amazon DeepAr, where accuracy was improved from 86.66 percent to 96.12 percent for Malaysian data and from 61.45 percent to 68.92 percent for Grenoble data for 24 hours ahead forecasting, from 61.5 percent to 68.9 percent for Malaysian data, and from 43.8 percent to 52.3 percent for Grenoble data for 1 month ahead forecasting.

On the prospective use of deep learning systems for earthquake forecasting over Schumann Resonances Signals

Carlos Cano, Ruxandra Stoean, Nuria Novas Castellano, Manuel Fernandez Ros, Gonzalo Joya Caparrós and Jose Antonio Gázquez Parra

Abstract: The relationship between Schumann Resonances and earthquakes was proposed more than 50 years ago. However, the experimental support has not been fully established. A considerable amount of recent studies have focused on the relationship between a single Earthquake and the Schumann Resonance signal variation around this earthquake, obtaining preliminary support for the existence of the link. Nonetheless, they all lack a systematic and general approach. In this research, we propose a novel methodology to detect the presence of relevant earthquakes based on the Schumann Resonance. The methodology is based on a Deep Learning Framework composed of a pretrained Variational Auto-Encoder followed by an LSTM network and a fully connected layer with a sigmoid output. The results reveal the uncovered relationship between Earthquake activity and Schumann Resonance signal using the novel methodology, being the first automatic earthquake detector based on Schumann Resonance signal.

Reconstructed Phase Spaces and LSTM Neural Network Ensemble Predictions

Sebastian Raubitzek and Thomas Neubauer

Abstract: We present a novel approach of combining the concept of reconstructed phase spaces with neural network time-series predictions. The presented methodology aims to reduce the parametrization problem of neural networks and improve autoregressive neural network time-series predictions. The idea is first to interpolate a data set based on its reconstructed phase space properties and then filter an ensemble prediction based on its phase space properties. The corresponding ensemble predictions are made using randomly parameterized LSTM (Long Short Term Memory) neural networks. These neural networks then produce a multitude of auto-regressive predictions, which are then filtered to achieve a smooth reconstructed phase space trajectory. Thus we can circumvent the prob- lem of parameterizing the neural network for each dataset individually. Here, the interpolation and the ensemble prediction aim to produce a smooth trajectory in a reconstructed phase space. The best results are compared to a single hidden layer LSTM neural network and benchmark results from the literature. The results show that the baseline predictions are outperformed for all three discussed data sets, and one of the benchmark results from the literature is bested by the presented approach.

An Open-Source and Reproducible Implementation of LSTM and GRU Networks for Time Series Forecasting

Gissel Velarde, Pedro Brañez, Alejandro Bueno, Rodrigo Heredia and Mateo Lopez-Ledezma

Abstract: This paper introduces an open-source and reproducible implementation of Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) Networks for time series forecasting. We evaluated LSTM and GRU

networks because of their performance reported in related work. We describe our method and its results on two datasets. The first dataset is the S&P BSE BANKEX, composed of stock time series (closing prices) of ten financial institutions. The second dataset, called Activities, comprises ten synthetic time series resembling weekly activities with five days of high activity and two days of low activity. We report Root Mean Squared Error (RMSE) between actual and predicted values, as well as Directional Accuracy (DA). We show that a single time series from a dataset can be used to adequately train the networks if the sequences in the dataset contain patterns that repeat, even with certain variation, and are properly processed. For 1-step ahead and 20-step ahead forecasts, LSTM and GRU networks significantly outperform a baseline on the Activities dataset. The baseline simply repeats the last available value. On the stock market dataset, the networks perform just like the baseline, possibly due to the nature of these series. We release the datasets used as well as the implementation with all experiments performed to enable future comparisons and to make our research reproducible.

Forecasting nonlinear causal relationships between crude oil prices and exchange rates using machine learning methods

Witold Orzeszko

Abstract: The relation between crude oil prices and exchange rates has always been of interest to academics and policy analysts. There are theoretical transmission channels which justify evidence for such relation. Based on the nonlinear Granger causality test we found strong bidirectional causal relations between crude oil prices and two currency pairs: EUR/USD, GBP/USD, and much weaker between crude oil prices and USD/JPY. We made an attempt to find an effective strategy to forecast crude oil prices using investigated exchange rates as regressors and vice versa. To this aim we applied selected machine learning methods of time series forecasting.

APUMPEDI: Approximating Pan Matrix Profiles of Time Series Under Unnormalized Euclidean Distances by Interpolation

Jing Zhang and Daniel Nikovski

Abstract: The Matrix Profile (MP) of a time series has been proposed as a versatile primitive for many data mining tasks. As a companion time series, the MP records distances between nearest neighbors of subsequences in the original time series. The Pan Matrix Profile (PMP) is a matrix with each row being an MP corresponding to a single subsequence length, and computing explicitly an exact PMP is slow. We propose an approximation algorithm called APUMPEDI to compute the PMP under the unnormalized Euclidean distance based on MP algorithms combined with interpolation. We validate their efficiency and effectiveness through extensive numerical experiments on both real-world and synthesized data sets.

Macroeconomic Predictions using Payments Data and Machine Learning

Ajit Desai and James Chapman

Abstract: Predicting the economy's short-term dynamics---a vital input to economic agents' decision-making process---often uses lagged indicators in linear models. This is typically sufficient during normal times but could prove inadequate during crisis periods such as COVID-19. This paper demonstrates: (a) that payments systems data which capture a variety of economic transactions can assist in estimating the state of the economy in real time and (b) that machine learning can provide a set of econometric tools to effectively handle a wide variety in payments data and capture sudden and large effects from a crisis. Further, we mitigate the interpretability and overfitting challenges of machine learning models by using the Shapley value-based approach to quantify the marginal contribution of payments data and by devising a novel cross-validation strategy tailored to macroeconomic prediction models.

Outliers impact on parameter estimation of Gaussian and non-Gaussian state space models: a simulation study

F. Catarina Pereira, A. Manuela Gonçalves and Marco Costa

Abstract: State space models are powerful and quite flexible tools that allow dealing with systems that vary significantly over time due to their formulation, because the models parameters vary over time. A major advantage of these models is the possibility of explicitly integrating the unobservable components of a time series by relating to each other stochastically. There are several books in the literature that describe state space models in detail [1], [2], [3], [4], [5]. State space models have in their structure a latent process, the state, which is not observed. To estimate it, the Kalman filter is typically used, as it is a recursive algorithm that, at each time, computes the optimal estimator, in the sense that it has the minimum mean squared error of the state, when the model is fully specified, and one-step ahead predictions by updating and improving the predictions of the state vector in real time when new observations become available.The Kalman filter was originally developed by control engineering in the 1960s in one of Kalman's papers [6] describing a recursive solution to the linear filter problem for discrete time. Today, this algorithm is applied in various areas of study. Usually, to estimate the unknown parameters of the model, the maximum likelihood method is used by assuming normality of the errors; however, this assumption cannot always be guaranteed. Non-parametric estimation methods can be a strong contribution when it comes to the initial values of iterative methods used to optimize the likelihood function, which often do not verify the convergence of the algorithms due to the initial choice of these parameters. For example, [7] propose estimators based on the generalized method of moments, the distribution-free estimators, where these estimators do not depend on the distribution of errors. Nevertheless, even if the assumption of normality of errors is not verified, the Kalman filter still returns optimal predictions within the class of all linear estimators. However, the optimal properties of Kalman filter predictors can only be ensured when all state space models parameters are known. When the unknown parameter vector is replaced by its estimate, the mean squared error of the estimators is underestimated. Outliers ate another important issue in time series modeling. Time series data are typically dependent on each other and the presence of outliers can impact parameter estimates, forecasting and also inference results [8]. In the presence of incomplete data and outliers in the observed data, [9] developed a modified robust Kalman filter. [10] showed that linear Gaussian state space models are suitable for estimating the unknown parameters and consequently affect the state predictions, especially when the measurement error was much larger than the stochasticity of the process. Therefore, to evaluate the quality of the estimates of the unknown parameters of the linear invariant state space model, this paper presents four simulation studies: the first is based on the linear Gaussian state space model; the second is based on the linear Gaussian state space model with contaminated observations; the third is based on the linear non-Gaussian state space model with exponential errors; and the last is based on the linear non-Gaussian state space model with exponential errors and contaminated observations. For each of the four studies, several scenarios were tested, in which 2000 samples with valid estimates, that is, estimates within the parameter space of size n (n=50,100,200,500), were simulated. The results obtained were evaluated in terms of the difference between the maximum likelihood estimate and the true value of the parameter and the rate of valid estimates.

Linear Programming-Based Bi-Objective Optimization for Short Univariate Time Series Forecasting

Santhosh Kumar Selvam, Chandrasekharan Rajendran and Ganesh Sankaralingam

Abstract: This paper proposes a Linear Programming-based time series Forecasting Al-gorithm – considering Multiple Accuracy Measures (LPFA-MAM) that helps forecast sub-annual short univariate time-series data of a retail e-commerce company. The company's main challenge is to get a more accurate daily revenue forecast at the company level while simultaneously reducing the opportunity cost and inventory cost that arises from either under-forecasting or over-forecasting the revenue. The proposed algorithm generates forecasts optimized for a pair of accuracy measures instead of just one measure. The algorithm is based on the ϵ - constraint-based multi-objective optimization method. The accuracy measure pairs used in this paper are Mean Absolute Error - Maximum Absolute Error (MAE-MaxAE) and the Mean Absolute Percentage Error - Maximum Absolute Percentage Error (MAPE-MaxAPE). We compare the performance of the LPFA-MAM with several industry-

standard forecasting methods using accuracy measures commonly reported in the literature for three forecast horizons, namely, long-term, medium-term, and short-term. For the short-time series studied in our paper, the LPFA-MAM performs the best in the long- and medium-term horizon. Across all forecast horizons, the LPFA-MAM has the least MaxAE and MaxAPE, reducing the over-and under-forecast errors. The LPFA-MAM can yield interpretable linear forecasting models and is quite flexible.

Exploring the performance of Test Statistics of Time Series Outlier Detection: Evidence from Robust Estimation Methods

Tanzeela Yaqoob and Arfa Maqsood

Abstract: Outliers in time series data are of immense importance in model estimation and forecasting. Understanding the importance of false detection of outliers in various Autoregressive process, the present study aims to evaluate the performance of test statistics by utilizing various models through robust estimation of errors. Numerous data generating process have been employed via simulation at different values of the estimates, location and size of outlier, sample sizes and classical cut-off to access the power of the test statistics for false detection of outlier type. The estimation of errors and its standard deviation has been accomplished through robust scale estimators and filtering procedures. The findings reveal that behavior of test statistic in detecting the type of outliers are influenced by location and size of outliers, parameter values and to some extent size of the series. The detection of duplicate outliers especially for TC at various transient values is controlled through robust estimation. On the whole the method of estimation of residual standard deviation influences the sampling behavior of the test statistics. The k-step within sample forecasting performance with and without outliers has judged by means of various diagnostic measures of prediction.

Combining ARIMA-SVR and Fuzzy Time Series Model based on Hedge Algebra to Forecast Bitcoin

Dinh Thuan Nguyen and Tung Hoang

Abstract: Some researchers show that the fuzzy time series (FTS) models usually give good forecasting results on time series with a small number of observations, while ARIMA (Auto-Regressive Integrated Moving Average) model or ARIMA combined with other models such as SVR or ANN or LSTM often gives good forecasting results on time series with a large number of observations. In this paper, we introduce how to combine FTS based on Hedge Algebras (FTSHA) and ARIMA-SVR to promote the strengths of these two types of models to create a new model with better forecasting ability. The Bitcoin time series is used for testing the new model's performance. Experimental results show that the new model gives better forecasting results than the existing ones

Big data oriented to innovation through quality techniques

Salvador Pérez-García and Manuel García-García

Abstract: The paper deals with the new capabilities of science associated with massive data analysis. To do this, first, the development of how these capacities have been developed and in what current state we find ourselves today will be presented. Then, the main challenges that are expected to obtain advances in data analytics techniques will be indicated, as well as the different statistical approaches that allow providing solutions to them. Afterwards, the most representative software and languages will be briefly presented, to finally present application cases in which big data is capable of providing improvements in quality, innovation and sustainability in both processes and products in different sectors of activity.

Time Series Clustering of High Gamma Dose Rate Incidents

Mohammed Al Saleh, Béatrice Finance, Yehia Taher, Ali Jaber and Roger Luff

Abstract: In this paper, we propose an unsupervised machine-learning-based framework to automate the process of extracting suspicious gamma dose rate incidents from the real unlabeled raw historical data measured in the

German Radiation Early Warning Network and identifying the underlying events behind each. This raises the research problem of clustering unlabeled time series data with varying lengths and scales. Based on the many evaluations, we demonstrate that the state of the art's most popular time-series clustering models were not suitable to perform this task. This motivated us to introduce our own approach.

A Hybrid Model of VAR-DCC-GARCH and Wavelet Analysis for Financial Markets Volatility

Maryam Nafisi Moghadam and Shahram Fattahi

Abstract: The purpose of this study is to investigate the time-varying co-movement between the volatility of gold, exchange rate, and stock market returns in Iran, using weekly data from September 27, 2013, to December 3, 2021. This paper deployed VAR-DCC-GARCH and continuous wavelet for correlation analysis. The results of VAR-DCC-GARCH model indicate that a positive and relatively high conditional correlation exists in the daily exchange rate and gold returns volatility. Furthermore, the conditional correlation is lower between the exchange rate - stock market returns and the daily gold -stock market returns. Then, we applied VAR-DCC-GARCH for volatility measurement in three indicates. This volatility is analyzed using CWT in the time-frequency (short-, medium- and long-run) frequencies. The results show that in the short and long term, there is no correlation between the exchange rate and the stock market volatilities as well as the gold and the stock markets volatilities, but in the medium term, the correlation between the two markets is confirmed.

Multiplicative Nonstationary Volatility Models with Exogenous Information

Cristina Amado and Timo Teräsvirta

Abstract: In this paper, we propose a multiplicative nonstationary volatility model allowing for nonlinear behaviour driven by exogenous information. The new model extends the time-varying GARCH model of Amado and Teräsvirta (2013, 2017) by including an additional stochastic variable to allow the conditional variance to change smoothly between regimes. Modelling strategies for the proposed model are developed, and they rely on Lagrange multiplier tests. The estimation of the model is simplified by employing maximisation by parts and the asymptotic properties of the proposed estimators are also studied. Finite-sample properties of these procedures and statistical tests are examined by simulation. An empirical application to agricultural commodity returns illustrates the functioning of the model in practice.

Unbiased estimation of autoregressive models based on the SIMEX algorithm

Wilfredo Palma, Susana Eyheramendy, Felipe Elorrieta and Esteban Barrientos

Abstract: In many practical cases, the commonly used methods for parameter estimation in the context of time series data such as the maximum likelihood estimator (MLE) and the minimum square error estimator (MSE) exhibit biases. In particular, this problem arises when the parameters are close to the unit root or the process has additive noise. In order to improve these parameter estimation procedures, this work develops a simulation-extrapolation (SIMEX) methodology in the context of time series analysis. We consider a standard class of autoregressive processes (AR(p)) observed at regular times. Further, we study this problem in a recently developed class of irregularly autoregressive processes (IAR). Additionally, we analyze the effects of additive noise in the processes. The performance of the proposed SIMEX algorithm is investigated by means of extensive simulations and real-life data illustrations. These results show that the proposed methodology improves substantially the quality of the estimates from MLE and MSE methods, in terms of both bias reduction and estimation precision, in the context of both the regular/irregular and additive noise problems.

Deep Learning Applied to Wind Power Forecasting: a Spatio-Temporal Approach

Rubén del Campo, Eloy Anguiano, Álvaro Romero and José R. Dorronsoro

Abstract: In recent years, the development of the electricity system towards a production system based on mainly renewable energies has enabled society to move towards a more efficient and ecologically sustainable way of life. However, the generation of some of these energies are really hard to store and their good

performance is inevitably linked to weather conditions. Given the great importance that this problem has gained in view of the large percentage that renewable energies already occupy in the energy mix, in recent decades a considerable amount of research has been carried out in the field of machine learning with the aim of forecasting the energy production of wind or photovoltaic farms with the greatest possible accuracy. Usually, models that receive data in tabular format are used to make such predictions. However, in this article we propose to solve the problem of predicting renewable energy (wind power in this case) like a spatio-temporal prediction problem as if it were an image or video analysis problem. Furthermore, thanks to the advances made both in the field of GPU computing, which make possible the execution of complex algorithms in reasonable times, and in the field of machine learning with the implementation of new and more efficient optimization algorithms, the state of the art has evolved to what is now known as deep learning. On the one hand, energy production, as well as wind speed, temperature and other weather variables provided by Numerical Weather Prediction models (NWP) are time series and their temporary patterns are important justifying the temporal treatment. On the other hand, NWP variables are provided in a regular grid format (in terms of latitude and longitude). Thus, the data are arranged as different meteorological variables in the form of a grid, justifying the spatial treatment as if it were a low-resolution image, where the meteorological points are treated as pixels. Due to the temporal and spatial nature of the prediction data from meteorological providers, we open the possibility of using models that have demonstrated high performance in this type of problems, and those are the models based on deep neural networks. Specifically, we propose the use of different architectures belonging to the field of deep learning based on convolutional networks (CNN), and/or multi-attention mechanisms based on Transformers such as the VisionTransformer (ViT) to encode the spatial information of meteorological data combined with recurrent networks (RNN) or LSTMs to capture the temporal information of the problem. Based on these architectures, we carry out implementations where we make combinations between them trying to provide the neural network with a greater ability to appreciate features that would be very complicated for other more classical implementations. Some of these networks have been used in previous works, but always individually and in most cases they have only been applied to similar problems such as weather prediction or other very different ones such as voice signals or video recognition. But we do not know of any case where they have been applied to the specific problem at hand. For this reason, the ultimate goal of this particular article is to carry out an initial benchmark that compares the performance measured between different types of architectures and their degradation over the prediction time horizons. The proposed architectures are: CNN, LSTM, LSTM+CNN (Stacked), LSTM+CNN (Parallel), ConvLSTM and ViT. In view of the good quality of the results obtained, it can be observed that these approaches open a great possibility of improvement of the chosen base models, thus setting a good precedent in the investigation of this problem. In particular, the most promising architecture has turned out to be the VisionTransformers (ViT), which suggests the use of models that have been quite successful in other domains such as text (NLP), image or video (computer vision).

Dynamical Component Analysis: Updated and improved algorithm, applications, and limitations

Monika Warmuth, Philipp Romberger, Knut Hüper and Christian Uhl

Abstract: We will present an updated version of the dynamical component analysis (DyCA) and show by simulations that DyCA not only leads to the low-dimensional decomposition of the data but also separates the deterministic part of the signal from stochastic components.

Traffic flows time series clustering in flooding risk areas using origin-destination mobile phone data

Maurizio Carpita, Giovanni De Luca, Rodolfo Metulini and Paola Zuccolotto

Abstract: It is well known that extreme weather events often have huge social consequences for communities and individuals. Their immediate effects consist of loss of human life, devastation of crops, damage to goods, and deterioration of overall health and wealth conditions. Considered their social and economic impact, the statistical study of extreme weather phenomena can also be approached from a management perspective. In fact, natural Disaster Management (Mishra et al., 2019) recommends the development of a framework of exposure risk that can exploited in an early warning perspective. In this work we focus attention on floods. To draw

flooding risk exposure maps is of fundamental importance, in order to face, in the best possible way, a flooding event. Such maps cannot ignore human presence and people mobility, but they traditionally assume a constant crowding over time. This assumption is far from reality, especially in metropolitan areas, so, a more detailed description of people presence and mobility is a critical issue to determine an accurate flooding risk. To do that, mobile phone network data have been used to obtain a dynamic monitoring of crowding in areas with hydrogeological criticality (Balistrocchi et al., 2020). Another approach consists of using mobile phone origin-destination signals, in order to recover information on traffic flows and then build statistical models, able to give accurate forecasts of people mobility. Metulini and Carpita (forthcoming) proposed a model, based on combining vector autoregressive with exogenous covariates and dynamic harmonic regression. They applied the method to the case study of Mandolossa (an urbanized area subject to flooding, located on the western outskirts of Brescia) using hourly data from September 2020 to August 2021. The method worked quite well, but residuals exhibited a leptokurtic distribution with heavy tails determined by a number of extreme events (i.e., days with particularly high or low traffic flows). In this talk we propose to use the method of time series clustering based on copula functions, proposed by De Luca and Zuccolotto (2011), in order to cluster the residuals time series with respect to their upper tail dependence. The aim of the analysis is to obtain clusters of areas for which extreme events (in terms of extremely high traffic flows) tend to occur together. REFERENCES Balistrocchi M., Metulini R., Carpita M., Ranzi R. (2020). Dynamic maps of human exposure to floods based on mobile phone data. Natural Hazards and Earth System Sciences, 20(12), 3485-3500. De Luca G., Zuccolotto P. (2011). A tail dependence-based dissimilarity measure for financial time series clustering. Advances in Data Analysis and Classification, 5(4), 323-340. Metulini R., Carpita M. (forthcoming). A time series approach for flooding risk outbreaks' early detection using mobile phone traffic flows data. Forthcoming. Mishra D., Kumar S., Hassini E. (2019). Current trends in disaster management simulation modelling research. Annals of Operations Research, 283(1), 1387-1411.

Modelling and predicting the dynamics of confirmed Covid-19 cases based on climate data

Yuzhi Cai, Fangzhou Huang and Jiao Song

Abstract: We examine the impact of climate variables on the transmissibility of coronavirus (Covid-19) and predict the dynamics of confirmed Covid-19 cases using time series models and climate data. Although our method can be used in other parts of the world, our research focuses on Wales, UK. We find that temperature changes have a significant negative impact on the transmissibility, implying that as winter approaches, the spread of Covid-19 is likely to be aggravated. However, our findings also show that, due to the particular climate condition in Wales with considerably less volatile seasonal temperature change, the economic impact of the climate variables on suppressing the spread of Covid-19 in Wales is limited, suggesting that effective measures and potential intervention are crucial, particularly, during another wave. We also find that the accuracy of our forecasting results is high, indicating that our models are robust for forecasting and the climate variables are significant predictors of the Covid-19 daily cases in Wales. Our findings provide numerical evidence and are of importance when inform evidence based policy making on public health in Wales.

Dynamic Asymmetric Causality Tests with an Application

Abdulnasser Hatemi-J

Abstract: Testing for causation—defined as the preceding impact of the past value(s) of one variable on the current value of another one when all other pertinent information is accounted for—is increasingly utilized in empirical research of the time-series data in different scientific disciplines. A relatively recent extension of this approach has been allowing for potential asymmetric impacts since it is harmonious with the way reality operates in many cases (Hatemi-J, 2012). The current paper maintains that it is also important to account for the potential change in the parameters when asymmetric causation tests are conducted, as there exists a number of reasons for changing the potential causal connection between the variables across time. The current paper extends therefore the static asymmetric causality tests by making them dynamic via the usage of subsamples. An application is also provided consistent with measurable definitions of economic or financial bad as well as good news and their potential causal interaction across time.

The implications of the ICT adoption for wages and employment

Pablo Casas and Concepción Román

Time series analysis of topological indicators of the network of connections in the insurance sector during the greatest financial crises of the 21st century.

Anna Denkowska and Stanisław Wanat

Abstract: The aim of the research is to identify the similarities and differences in the network of interconnections between the world's largest insurance institutions during global financial crises. We assess the dependencies in the insurance sector in the context of the possibility of contagion with negative effects of unwanted financial events, which may be of systemic importance. In the empirical analysis, we combine taxonomic tools with network theory tools. The constructed networks of connections, obtained in various ways with the use of various metrics, make it possible to identify specific network behaviors in the context of the assessment of the Systemic Risk generated by the insurance sector. When assessing the structure of the network, we take into account the topological indicators of the entire network as well as indicators assessing individual insurance units. We examine the periods of the subprime mortgage crisis, excessive public debt in Europe, the immigration crisis, the Covid 19 crisis, and the war crisis in Ukraine.

Renewable energy transition in Europe: a multi-national analysis.

Andrea Savio and Mariangela Guidolin

Abstract: This paper proposes an investigation of the renewable energy transition in Europe through a multivariate diffusion model based on ordinary differential equations. Specifically, the model is applied to 18 European countries to analyze the dynamic interdependencies between renewables and a couple of competitors in order to identify similar social, economic, and geographical factors that enable or hinder renewables exploitation. This cross-country assessment can serve as a helpful starting point for national forecasting purposes and the design of targeted policy interventions that can stimulate or accelerate broader adoption of green energy sources and internalize the energy production chain.

Utilising Empirical Mode Decomposition, CNN-LSTM Network with Bayesian Optimisation for Accurate Short-term Residential Load Forecasting

Ashkan Lotfipoor, Sandhya Patidar and David P. Jenkins

Abstract: Near accurate residential load, forecasting has become a non-trivial requirement for developing effective management and planning strategies/policies for a resilient energy system. However, due to the stochastic nature of energy load, the energy load is hard to predict accurately and error-prone. Because time series data can display a wide range of patterns, it is often helpful to break it down into numerous components, each representing a different underlying pattern category. In this paper, a deep-learning model with a Convolutional neural network (CNN) and long short-term memory (LSTM) based on empirical mode decomposition (EMD) and feature engineering is proposed to forecast short-term residential load. In addition, for the proposed deep learning model, the Bayesian optimisation approach is used to find the best combination of hyperparameters. The proposed scheme uses a feature decomposition technique to improve the model's accuracy. The load profile is decomposed into intrinsic mode functions by empirical mode decomposition to capture periodic features. Then, a CNN-LSTM neural network is trained by three-dimensional input data. Finally, the prediction results of all intrinsic mode functions are combined with an advanced percentile-based bias correction to determine an aggregated output for load forecasting. The results are validated using the developed model on real-world data from a household in Fintry, Scotland. The performance of the introduced architecture is evaluated by comparing its results with those of standard methods, random forest, gradient boosting decision trees, and an LSTM network. The experimental results demonstrate that the prediction accuracy of the proposed approach can be significantly improved by utilising EMD and feature engineering.

The experiments also showed that the recommended number of intrinsic mode functions (IMF) is seven based on both prediction accuracy and running time.

Demand forecasting under lost-sales stock policies with Tobit Exponential Smoothing

Diego J. Pedregal, Juan R. Trapero and Enrique Holgado

Abstract: Demand forecasting is a crucial task within supply chain management. Stock control policies are directly affected by the precision of the demand probabilistic forecasts. For instance, safety stocks and reorder points are based on those forecasts. However, typically, forecasting and replenishment policies have been studied separately. In this work, we explore the influence of inventory assumptions on the selection of the forecasting model. Particularly, when the stock policy follows a lost-sales context and the demand is estimated by means of sales data. In that case, the forecasting models should deal with censored demand estimations. The literature about forecasting demand under lost-sales is scarce. Previous research may be divided in two groups depending on how the unobserved part of the demand distribution is estimated: i) those that try to estimate real demand using alternative sources of new information or changing the information acquisition process (see e.g., \cite{wecker1978}, \cite{Sachs2014}, \cite{Heese2010} and \cite{Stefanescu2011}); and ii) those that try to estimate real demand using only sales information (e.g., \cite{Nahmias1994}, \cite{Agrawal1996} and \cite{Kaplan1958}). In this work we focus on a solution along the second strand of literature, based on the development of censored dynamic models. The treatment of censored lost sales is tightly related to the censored demand forecasting problems (type I censoring). James Tobin in its seminal work \cite{Tobin1958} analysed censored data by means of a regression model for household expenditures. Later, this analysis was generalised in several ways until arriving at dynamic state space models estimated by a Tobit Kalman filter (see \cite{Allik2014} and \cite{Allik2016}). This solution seems quite convenient when dealing with time varying censoring levels, that is the usual situation in real life, while most papers only consider constant censoring levels. As a matter of fact, the censoring level depends on the forecasts themselves. This limitation is overcome elegantly by the Tobit Kalman filter \cite{Allik2014}. An additional step common in demand forecasting is the use of Exponential Smoothing models to deal with trend and seasonal components. Such models (as well as ARIMA) may be cast in a state space framework in innovations form, meaning that the states may be observed directly from the data (see equation (\ref{eq:lssrec}) below), with a considerably reduction of computation burden with respect to standard state space models and standard Kalman filters. The combination of all the previous elements led to the development of either Tobit Exponential Smoothing models or Tobit ARIMA models for censored demand forecasting. These models are able to deal with time varying censoring levels as well. To the best of authors' knowledge, it is the first time that the Tobit Kalman Filter is applied to supply chain demand forecasting and it may represent a game changer when dealing with a lost-sales context. Consider the following linear Gaussian innovations state space model \begin{equation} \label{eq:sysinnLinear} \begin{array}{c} x_{t} = F x_{t-1} + g \epsilon_{t},\\ y_t = w x_{t-1} + \epsilon_t, \end{array} \end{equation} \noindent where y_t is the output time series; x_t is a state vector; ϵ_t is a scalar perturbation such that $\epsilon_t \sim N(0,\sigma^2)$ and F, w and g are system matrices of appropriate dimensions (g is the so called Kalman gain). Given an initial state vector x_0, the state vector at each time stamp can be estimated recursively from the above equations without any uncertainty making the estimation of the states rather easy, i.e., \begin{equation} \label{eq:lssrec} \begin{array}{c} \epsilon_t=y_t-w x_{t-1}, \\ x_{t}=F x_{t-1}+ g \epsilon_t\\ \end{array} \end{equation} One trouble is that the initial state is crucial to estimation, especially in small samples, and it is generally not known. To get round this problem several solutions have been proposed in the literature. The initial state vector may be treated as a deterministic quantity and estimated along with the rest of parameters by maximum likelihood or by using an information filter as in \cite{Hyndman2008}. However, another approach has been widely used in the literature, consisting of considering the initial state vector as a multivariate stochastic variable and estimating it by diffuse initialization, see e.g., \cite{koopman2012} or the augmented Kalman filter, proposed by \cite{deJong1991}. The linear Gaussian innovations state space model with Tobit censoring from above, where y_t^* is the uncensored variable and y_t is its censored counterpart is \begin{equation} \label{eq:sys} \begin{array}{l} y_t^* = w x_{t-1}+ \epsilon_t^*\\ x_{t} = Fx_{t-1}+ g \epsilon_{t}^*\\ y_t= \left\{ \begin{array}{ll} y_{t}^*, & y_{t}^* < Y_{max,t} \\ Y_{max,t} & y_{t}^* \geq Y_{max,t} \end{array} \right. \end{array} \end{equation} \noindent and $\epsilon_t^* \sim N(0,\sigma^2)$. If the uncensored y_t^* variable were observed, given a

known initial value for the states x_1 the rest of states would be known exactly and the one step ahead forecast would be $\textrm{E} (y_{t}^*|x_{t-1}) = w x_{t-1}$ with variance $\textrm{Var} (y_{t}^*|x_{t-1}) = \sigma^2$, as in (\ref{eq:lssrec}). But y_t^* is not observed at some time stamps and then the challenge becomes to estimate the moments of the distribution of y_t^* based on the moments of the censored data y_t only. In this work we show that the recursions for the Tobit model become \begin{equation} \begin{array}{rl} \textrm{E} (y_{t}|x_{t-1}) = &P_{un,t} (w x_{t-1} - \sigma m_t) + P_{max,t} Y_{max,t}\\ \epsilon_t = & y_t - \textrm{E} (y_{t}|x_{t-1}) \\ \textrm{E}(x_t | Y_{t})= &F x_{t-1} + P_{un,t} / (1 + c_t / P_{un,t} - m_t^2) g \epsilon_t, \end{array} \end{equation} \noindent where $P_{un,t}$ is the probability of observation at time t to be uncensored, $P_{max,t}$ is the probability of an observation to be censored from above, m_t is the inverse Mill ratio at time t, and c_t is another parameter involved in censored distributions. Estimation of model parameters is carried out by maximum likelihood including the initial state vector as known fixed values. Taking ϕ as the Gaussian standard pdf, the likelihood function for the Tobit model is \begin{equation} \label{eq:sys2b} \begin{array}{c} L(y_1,\dots,y_n) = \prod_{y_t=Y_{max,t}} P_{max,t} \times \prod_{y_t < Y_{max,t}} \frac{1}{\sigma} \phi \left(\frac{y_t - w x_{t-1}}{\sigma} \right). \\ \end{array} \end{equation} In this work we compare the Tobit Kalman filter applied to censored demand with other standard ad-hoc solutions, either in simple cases where demand is assumed random and more complex cases involving trends and seasonal components. The examples illustrate the significant advantages that brings the proposed model in terms of costs and service level.

Online estimation methods for irregular autoregressive models

Felipe Elorrieta, Lucas Osses, Matias Caceres, Susana Eyheramendy, Wilfredo Palma and Esteban Barrientos

Abstract: In the last decades, due to the huge technological growth observed, it has become increasingly common that a collection of temporal data rapidly accumulates in vast amounts. This provides an opportunity for extracting valuable information through the estimation of increasingly precise models. But at the same time it imposes the challenge of continuously updating the models as new data becomes available. Currently available methods for addressing this problem, the so-called online learning methods, use current parameter estimations and novel data to update the estimators avoiding using the full raw data. In this work we consider three online learning algorithms for parameters estimation in the context of time series models. In particular, the methods implemented are: gradient descent, Newton-step and Kalman filter recursions. These algorithms are applied in the recently developed irregularly observed autoregressive (iAR) model. The estimation accuracy of the proposed methods is assessed by Monte Carlo experiments. The results obtained show that the proposed online estimation methods allow to estimate precisely the parameters that generate the data both for the regularly and irregularly observed time series and in less computational time. Moreover, we show that the proposed methods are able to adapt the parameter estimates quickly when the time series behavior changes, unlike batch estimation methods.

Noise inference for ergodic Lévy driven SDE

Hiroki Masuda, Lorenzo Mercuri and Yuma Uehara

Abstract: We study inference for the driving Lévy noise of an ergodic stochastic differential equation (SDE) model, when the process is observed at high-frequency and long time and when the drift and scale coefficients contain finite-dimensional unknown parameters. By making use of the Gaussian quasi-likelihood function for the coefficients, we derive a stochastic expansion for a functional of the unit-time residuals, which clarifies some quantitative effect of plugging in the estimators of the coefficients, thereby enabling us to take several inference procedures for the driving-noise characteristics into account. We also present new classes and methods available in YUIMA for the simulation and the estimation of a Lévy SDE model. We highlight the flexibility of these new advances in YUIMA using simulated and real data.

CARMA(p, q) Hawkes processes

Lorenzo Mercuri, Andrea Perchiazzo and Edit Rroji

Abstract: The insurance claims are modeled through an Hawkes process with random jump size where the intensity follows a CARMA(p,q) model with a self-exciting effect due to the presence of the counting process as a driving noise. Recent empirical studies have showed that the autocorrelation function of the number of claims has a decreasing trend but it is not strictly monotonic. A standard Hawkes model with exponential kernel for insurance claims developed in Swishchuk et al. (2021) is able to reproduce a exponential monotonic decreasing behaviour of the autocorrelation function. Our model, that can be seen as a generalization of the Hawkes process with exponential kernel, is able to generate a more flexible shape for the autocorrelation function and obtain a more realistic dependence structure observed in the number of claims.

Synthetic subject generation with coupled coherent time series data

Xabat Larrea, Mikel Hernandez, Gorka Epelde, Andoni Beristain, Cristina Molina, Ane Alberdi, Debbie Rankin, Panagiotis Bamidis and Evdokimos Konstantinidis

Abstract: A huge amount of health and well-being data is collected daily, but little of it reaches its research potential because personal data privacy needs to be protected as an individual's right, as reflected in data protection regulation. Moreover, the data that does reach the public domain will typically have undergone anonymization, a process that can result in a loss of information and, consequently, research potential. Lately, synthetic data generation, which mimics the statistics and patterns of the original, real data on which it is based, has been presented as an alternative to data anonymization. As the data collected from health and well-being activities often has a temporal nature, this data tends to be time series data. The synthetic generation of this type of data has already been analyzed in different studies. However, in the healthcare context, time series data has a reduced research potential without the subjects' metadata which is essential to explain the temporal data. Therefore, in this work, the option to generate synthetic subjects using both time series data and subject metadata has been analyzed. Two approaches for generating synthetic subjects are proposed. Real time series data is used in the first approach, while in the second approach, time series data is synthetically generated. Furthermore, the first proposed approach is implemented and evaluated. The generation of synthetic subjects with real time series data has been demonstrated to be functional, whilst the generation of synthetic subjects with synthetic time series data requires further improvements to demonstrate its viability.

Renewable Energy Installations and Policy Assessment: Evidence from cointegration and non-causality in Greece (1987-2020)

Georgios Evangelidis, Konstantinos Konstantakis and Panayotis G. Michaelides

Abstract: The ongoing energy crisis in Europe has deteriorated due to the Russian invasion in Ukraine that led most European governments to stop their natural gas supply from Russia, as a protest against the war. In this context, the dire need for energy substitutes throughout the EU economies is on the agenda of most policy makers. The installation and growth of Clean and Renewable Energy remains until today a strong substitute against the dependence of fossil fuels. Nonetheless, due to the fact that the vast majority of renewable energy investments depend critically on private funds, the assessment of policies that could help in increasing these investments is of outmost importance. There is a vast literature on the connection between energy consumption and economic growth. However, most of the studies can be classified based on their primary hypothesis tested, under four main categories. This, is possible, due to the fact that, based on the relevant literature, the energy consumption and economic growth relation can be analyzed under four hypotheses. Analytically, the hypotheses are the following: 1. The growth hypothesis assumes energy as a major source of input into the growth process, and unidirectional causality exists from energy consumption to economic growth. In this scenario, energy conservation policies will have a negative impact on economic growth. 2. The conservation hypothesis implies that economic growth causes consumption of energy. Under this situation, conservation policy will not affect economic growth. 3. The feedback hypothesis implies a bi-directional relationship between energy consumption and economic growth. This hypothesis suggests any change in energy consumption will affect economic growth with a reverse effect. 4. The neutrality hypothesis indicates that energy consumption and economic growth are independent and do not affect each other. In this context, observing the causal relationship between electricity

consumption and economic growth provides a basis for understanding the level of maturity of countries and provides insight and direction for the design and implementation of environmental and energy policies. It is not surprising that the empirical results have yielded mixed results in terms of the four hypotheses (growth, conservation, neutrality, and feedback) related to the causal relationship between electricity consumption and economic growth. Of course, a wealth of econometric techniques has been applied throughout the years in order to test one or more of the aforementioned hypotheses, like Fully modified OLS (FMOLS), Dynamic Panel Analysis, Autoregressive-Distributed Lags Models (ARDL), Autoregressive Vector Cointegration (VEC), Granger non-causality, Toda-Yamamoto non-causality etc. The choice of each method depended on the both the application sample, i.e. Time Series Panel Data sample; and the hypotheses that needed to be tested. In this study, besides the usual indicators we introduce a new approach for the investigation of the connection of Economic Growth and New Renewable Installations in Greece. New Renewable Installations is an urgent need for Energy adequacy in Europe and the way to promote them is a quest for clean energy. The current analysis casts light to the ways to increase new installations of Renewables. Using data for the Greek economy that span the period 1987 - 2020, we assess the impact of renewable energy production on the growth of the Greek economy, evaluating at the same time the various governmental policies implemented. To do so, we make use of non-causality tests and we unveil the cointegrating long-run relationships among key macroeconomic indicators like GDP, Trade, Renewable Energy Production, Renewable Energy Capacity and Gross Fixed Capital Formation. Based on our analysis, there is evidence in favor of (i) a long-run relationship among energy consumption and economic growth, and (ii) a rejection of the non-causality hypothesis between GDP and new renewable installations, as well as between Gross Fixed Capital Formation and new renewable installations. Therefore, policies that tend to incentivise new renewable installations should continue to be implemented focusing on carefully selected macroeconomic variables.

Forecasting Temperature of Satellites in Low Earth Orbit

Kiranmai Madisa, Sharmistha Dey and Paulo de Souza

Abstract: Today hundreds of satellites operate in Low Earth Orbit (LEO)[1]. Satellites in LEO are orbiting at an altitude ranging from 300km to 1500km[2] with an approximately 90min revolution. Therefore, an LEO satellite orbits the Earth 16 times per day. Each 90min orbit constitutes 60min of exposure to the Sun and 30min of eclipse and penumbra. Because of these regular orbits, one should expect the temperature of satellite components (electronics) to be predictable. However, electronics in space degrade faster than on Earth. The intense radiation and extreme temperature ranges in space affect electronics performance, affecting data transmission[3]. Data series can also have gaps resulting from faulty electronics, memory (read/write) failures, and downlink/uplink issues. Furthermore, sensors deteriorate over time. It is challenging to recalibrate or perform maintenance in biased sensors operating in satellites. Space qualified electronics can extend satellite hardware's useful life, yet sensors could produce noisy data. Noise can be a result of the temperature of operation, cosmic radiation and ageing. Hence temperature sensors in the satellite can produce results with higher noise as the satellite ages. Consequently, the satellite data will become unreliable over time. As a result, measures to regulate the temperature of electronic components in the satellite are less effective in older satellites[4]. This project addresses this problem by using the Autoregressive integrated moving average (ARIMA) model. The following is one of the many scenarios that illustrate the importance of time-series forecasting: Suppose that the satellite is in the middle of an eclipse with a circuit board at a temperature of -650C. This circuit board is intended to be used for an earth observation task in 26min. If the operational temperature range of this circuit board is from 400C to -100C, one should expect that this circuit board will be damaged or will not work properly if temperatures are outside the given range. Therefore, will it be possible to operate the circuit board in 26min? If not, will it need to be cooled down or warmed up? If so, when should the process of heating/cooling it down start? This project aims to address the above questions using the Autoregressive integrated moving average (ARIMA) model. Understanding the thermal dynamics of the satellite will assist in providing an answer to the above questions. Time series data will be collected, and time series analysis and forecasting will be conducted using the ARIMA model. It is proposed that this project will generate an ARIMA model for forecasting the thermal behaviour of the satellite that might possibly aid in managing the satellite resources effectively. References: 1. McDowell, J.C., The Low Earth Orbit Satellite Population and Impacts of the SpaceX Starlink Constellation. Astrophysical Journal Letters, 2020. 892(2). 2. Allahvirdi-Zadeh, A., K. Wang, and A. El-Mowafy, Precise Orbit Determination of LEO Satellites Based on

Undifferenced GNSS Observations. Journal of Surveying Engineering, 2022. 148(1). 3. Garzon, A. and Y.A. Villanueva, Thermal Analysis of Satellite Libertad 2: a Guide to CubeSat Temperature Prediction. Journal of Aerospace Technology and Management, 2018. 10. 4. von Lukowicz, M., et al., Characterisation of TEGs Under Extreme Environments and Integration Efforts Onto Satellites. Journal of Electronic Materials, 2015. 44(1): p. 362-370.

Hybrid K-Mean Clustering and Markov Chain for Network Accessibility and Retainability Prediction

Amel Salem Omer, Tesfaye Addisie Yemer and Dereje Hailemariam Woldegebreal

Abstract: Mobile network operators (MNOs) gather and analyze key performance indicators to continuously monitor the performance of their radio access net-works (RANs) in order to provide reliable services (KPIs). Accessibility, retainability, mobility, integrity, and availability are all KPIs that are clustered to measure network performance. Tracking KPIs help network managers to identify faults or anomalous events that can significantly degrade quality of service delivery and end-users experience. Taking corrective actions based on monitored parameters is nevertheless a reactive approach and contributes to network and service degradations until corrective actions are taken. In this work, we propose Markov chain as a systematic way of predicting, in probabilistic sense, network accessibility and retainability status, two important RAN performance parameters. 1530 real time KPI data is collected from an operator network for four months in hourly basis & using K-mean clustering approach the dataset are grouped into 6 clusters to build a Markov chain KPI status prediction models. The results demonstrate that the proposed models can achieve 94.61% prediction accuracy. This is a cost effective approach as the data is already available at the operator's network management system (NMS) and can be collected on a 24/7 basis.

Air pollution modeling for southern Poland through aggregated land-use regression

Katarzyna Kaczmarek-Majer, Olgierd Hryniewicz, Julia Soloch, Krzysztof Skotak and Anna Degórska

Abstract: In this work, we provide a statistical procedure to develop aggregated land-use regression models enabling annual point predictions of two health-related air pollutants, namely PM10 (Particulate Matter < 10 ug/m3) and NO2 (Nitrogen Dioxide). In experiments, validation of the proposed models and main findings from the feature selection analysis for various sources and grid cell sizes are described. Experiments suggest that the proposed models although simple, are efficient and interpretable for the prediction of annual air pollutants concentrations.

Value-at-Risk Forecasting with Hierarchical Outer Power Archimedean Copulas

Jan Gorecki

Abstract: Archimedean copulas (ACs) are dependence models frequently used in finance, insurance and risk management, e.g., for stress testing. In contrast to the most popular Gaussian and t-copulas, ACs allow for asymmetry in the joint tails, which is of particular interest, e.g., in risk management. In the talk, outer power (OP) transformations of Archimedean generators, which increase the modeling flexibility and statistical fitting capabilities of ACs restricted to a single parameter, will be recalled and their efficiency for Value-at-Risk (VaR) estimation based on time series for two portfolios of publicly traded stocks will be presented. The presented results will show that the OP transformation is able to improve the statistical fit of exchangeable Archimedean copulas, particularly of those that cannot capture upper tail dependence or strong concordance, as well as the statistical fit of hierarchical Archimedean copulas, especially in terms of tail dependence and higher dimensions. Given how comparably simple it is to include OP transformations into existing exchangeable and hierarchical Archimedean copula models, OP transformations provide a desired trade-off between computational effort and statistical improvement. Particularly, two different datasets of daily stock prices will be considered: the first one contains the five time series of stock prices of ADI (Analog Devices, Inc.), AVB (Avalonbay Communities Inc.), EQR (Equity Residential), LLY (Eli Lilly and Company) and TXN (Texas Instruments Inc.). These companies are selected as there exists a particular clustering of them given by their industry sector: ADI and

TXN belong to the high-tech industry, AVB and EQR to the real estate industry and LLY to the pharmaceutical industry. This clustering is then expected in the structure of the hierarchical copula models. The second dataset contains the first 10 time series of daily stock prices from the S\&P500 according to their market capitalization. For both datasets, the time span 2002-02-01--2019-02-01 is used. The presented study considers 19 copula models: Archimedean copulas (ACs), outer power ACs, hierarchical ACs and hierarchical outer power ACs for the popular families of Ali-Mikhail-Haq, Clayton, Frank and Joe, and then the standard benchmarks including Gaussian copulas, t-copulas and R-Vine copulas, as well as the quantile-based historical estimator, which is computed directly on the true profit-and-loss variates without any underlying time-series model. The presented results will clearly demonstrate that, in terms of the \emph{VaR violation ratio}, the outer power ACs and hierarchical outer power ACs estimators outperform all the remaining estimators in almost all cases, almost independently of the type of the generator family. Moreover, this implies that the outer power transformation consistently improves the non-outer power (hierarchical)AC estimators.

A Collection of Wisdom in Predicting Sector Returns

Hsiu-Lang Chen and Jolana Stejskalová

Abstract: This study investigates whether the aggregate investor information demand via Google search volume index (SVI) on all stocks in a sector can predict the sector performance. This study shows a sector’s abnormal SVI can predict the sector performance next month even afer controlling for the sector’s contemporaneous standardized earnings surprises and lagged returns on the market as well as the sector. Furthermore, this study shows a partial reversal in the sector‘s long-run performance, which is not completely consistent with the price pressure hypothesis. It indicates that parts of fundamental information about the sector can be captured by the sector‘s abnormal SVI on a more timely basis.

Relationship between stationarity and dynamic convergence of time series

Gerardo Covarrubias and Xuedong Liu

Abstract: The objective of this work is to analyze the consistency between stationarity and dynamic convergence of time series, since frequently, in econometrics only the issue of stationarity is addressed and the study of the reference dynamic equilibrium is omitted. Theoretically, both approaches are widely related and their study has the common purpose of making forecasts; however, from a methodological point of view, they are studied in a mutually exclusive manner. Therefore, in the face of an inappropriate specification of a model or the possible presence of spuriousness, the theoretical correspondence between these two properties could be incongruous in practice, especially for autoregressive model because we suppose ergodicity assumption.

The Vector Error Correction Index Model: Representation and Statistical Inference

Gianluca Cubadda and Marco Mazzali

Inflation Expectations, Inflation Target Credibility and the COVID-19 Pandemic: New Evidence from Germany

Dieter Nautz

Abstract: Using the exact wording of the ECB's definition of price-stability, we started a representative online survey of German citizens in January 2019 that is designed to measure long-term inflation expectations and the credibility of the inflation target. Our results indicate that credibility has decreased in our sample period, particularly in the course of the deep recession implied by the COVID-19 pandemic. Interestingly, even though inflation rates in Germany have been clearly below 2% for several years, credibility has declined mainly because Germans increasingly expect that inflation will be much higher than 2% over the medium term. We investigate how inflation expectations and the impact of the pandemic depend on personal characteristics including age, gender, education, income, and political attitude.

Online Classification of High Gamma Dose Rate Incidents

Mohammed Al Saleh, Béatrice Finance, Yehia Taher, Ali Jaber and Roger Luff

Abstract: In this paper, we propose a new method for choosing the most suitable time series classification method that can be applied to the online gamma dose rate incidents. We referred to the historical incidents measured in the German Radiation Early Warning Network and clustered them into several classes before testing existing classification methods. This raises the research problem of online classification of time series data with varying scales and lengths. Referring to the state of the art methods, we found that no specific classification method can fit our data all the time. This motivated us to introduce our own approach.

Multichoice Entropy Clustering for Time Series and Seasonality

Miguel Ángel Ruiz Reina

Abstract: Information Theory provides unsuspected elements of analysis for time series with seasonal behaviour. This new automatic clustering system provides a tool to classify decision-making by economic agents. The empirical application is in the tourist accommodation market, where international tourists must choose multiple accommodation options (hotels, tourist apartments, campsites and rural apartments). In contexts of seasonal uncertainty, it can resolve information gaps to understand human behaviour. The three-dimensional information of spatial extension, Spatial Location and Temporal extension is offered for the Spanish tourist market of foreigners who visit the Spanish autonomous communities between January 2001 and January 2022.

Do News Sentiments Predict Stock Price Direction?

Fennee Chong and Bharanidharan Shanmugam

Abstract: This paper explores whether news sentiments predict price directions of stocks listed on the S&P 500. To achieve the objective of this study, variables such as historical prices, technical indicators, business and political news sentiments ranging from 2015-2020 were tested using machine learning algorithms - Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB). Results indicated that news sentiments predict stock price directions whereby political news was confirmed to have a greater influence on stock prices than business news. Between the two machine algorithms used, XGB reports a much higher accuracy than SVM for all the models tested.

Combination of post-processing methods to improve high resolution solar irradiance forecasts in French Guiana

Rafael Alvarenga, Hubert Herbaux and Laurent Linguet

Abstract: Efforts have been made to improve Numerical Weather Prediction (NWP) forecasts with post-processing techniques, relying on statistical models to refine the weather forecasts. Most of the approaches used in literature suffer from two main deficiencies when applied to high-resolution data: 1) they require models with high capacity to retain nonlinear data fluctuations; 2) some are known to reduce the mean random error, however they may still generate subsequent biased forecasts. In this paper, three methods from three different approchaes are compared to improve 10-minute resolution NWP solar irradiance forecasts, namely a neural network as a model output statistics method, Kalman filter and kernel conditional density estimation. The results show that none of the methods, if used individually, improve the mean absolute error and mean bias error jointly. The neural netowrk is better on reducing the mean random error but induces systematic bias in original forecasts. Inversely, the Kalman filter is more able to reduce the mean bias error but fails on improving the mean random error considerably. The KCDE doesn't show expressive results in any of the metrics analysed. Furthermore, a combination of an neural network followed by Kalman filter post-processing results in major improvements both in the mean random error and the systematic bias of original forecasts, reducing the respective metrics by 44% and 87%.

Time Series sampling

Florian Combes, Ricardo Fraiman and Badih Ghattas

Abstract: Some complex models are frequently employed to describe physical and mechanical phenomena. In this setting we have an input X which is a time series, and an output Y = f(X) where f is a very complicated function, whose computational cost for every new input is very high. We are given two sets of observations of X, S1 and S2 of different sizes such that only f(S1) is available. We tackle the problem of selecting a subsample S3 ∈ S2 of smaller size on which to run the complex model f, and such that distribution of f(S3) is close to that of f(S1). We suggest five algorithms to solve this problem and show their efficiency using time serie data.

Measuring the impact of monetary policy on the stock market via an identified multivariate GARCH model

H. Herwartz, J. Roestel and F. Xu

Point and probabilistic forecast reconciliation for general linearly constrained multiple time series

Daniele Girolimetto and Tommaso Di Fonzo

Abstract: Time series can often be naturally disaggregated by various nested and/or crossed attributes of interest (Hyndman and Athanasopoulos, 2021, ch. 11). As an example, sales data can be disaggregated by product categories, and then by product subcategories, down to Stock Keeping Unit (SKU). Alternatively, sales data can be disaggregated by geographic divisions. Following Panagiotelis et al. (2020a), a hierarchical/grouped time series is a linearly constrained multiple time series consisting of a collection of time series that follows one or more hierarchical aggregation structure. Hierarchical forecast reconciliation is the post-forecasting process aimed to revise a set of incoherent base forecasts into coherent forecasts in line with cross-sectional/temporal/cross-temporal data structure. In both theoretical and empirical frameworks, most of the point and probabilistic hierarchical forecast reconciliation results move from the classic reconciliation formula valid for the structural representation of a hierarchical time series (firstly shown by Athanasopoulos et al., 2009). However, this formula holds for genuine hierarchical/grouped time series, sharing both the top and the bottom level variables. When a general linearly constrained multiple time series is considered, the projection approach reconciliation formula (van Erven and Cugliari, 2015, Di Fonzo and Girolimetto, 2021) gives a general solution. While it is well known that the classic structural reconciliation formula is equivalent to its projection approach counterpart, we show how a structural-like reconciliation formula may be derived for a general linearly constrained multiple time series. To do this, we describe how constrained multiple time series may be represented by a small number of basic variables that take the role of the bottom variables in the specific hierarchical/grouped framework. These tools would permit us to extend definitions, theorems and results found by Panagiotelis et al. (2020b) for probabilistic forecast reconciliation in a rather straightforward manner. To apply the theoretical framework, we propose a complete reconciliation procedure of probabilistic GDP forecasts, resulting in GDP forecasts coherent with both Income and Expenditure sides' forecasted series ('one number forecast'). The forecasting performance is evaluated on the Australian quarterly GDP series, as compared to the original proposal by Athanasopoulos et al. (2020). Our approach is applied within the same forecasting experiment, that considers ARIMA base forecasts for each time series from 1 quarter ahead up to 4 quarters ahead, using an expanding window. The base forecasts are reconciled using the R-package FoReco (Girolimetto and Di Fonzo, 2021). In general, the simultaneously reconciled probabilistic forecasts give results as good as those of Athanasopoulos et al. (2020). In addition, the newly proposed approach produces forecasts that are fully coherent with all economic constraints coming from National Accounts relationships, avoiding annoying discrepancies between the GDP forecasts from either Income or Expenditure sides. References Athanasopoulos, G., Ahmed, R.A., Hyndman, R.J. (2009), Hierarchical forecasts for Australian domestic tourism, International Journal of Forecasting 25, 146–166. doi:10.1016/j.ijforecast.2008.07.004. Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R.J., Affan, M. (2020), Hierarchical Forecasting, in Fuleky, P. (ed.), Macroeconomic Forecasting in the Era of Big Data, Cham, Springer, 689–719. Di Fonzo, T., Girolimetto, D. (2021), Cross-temporal forecast reconciliation: Optimal combination method and

heuristic alternatives, International Journal of Forecasting (in press). doi:10.1016/j.ijforecast.2021.08.004. Girolimetto, D., Di Fonzo, T. (2022). FoReco: Point Forecast Reconciliation. https://cran.r-project.org/package=FoReco Hyndman, R.J., Athanasopoulos, G. (2021), Forecasting: principles and practice (3rd ed.), Melbourne, OTexts. Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., Hyndman, R.J. (2020a). Forecast reconciliation: A geometric view with new insights on bias correction, International Journal of Forecasting 37, 343-359. doi:10.1016/j.ijforecast.2020.06.004. Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., Hyndman, R.J. (2020b), Probabilistic forecast reconciliation: properties, evaluation and score optimisation, Monash University, Department of Econometrics and Business Statistics, Working Paper 26/20. van Erven, T., Cugliari, J. (2015), Game-theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts, in Antoniadis, A., Poggi, J.M., Brossat, X. (eds.), Modeling and Stochastic Learning for Forecasting in High Dimensions, Berlin, Springer, 297-317.

The state of Africa's air transport market amid COVID-19, and forecasts for recovery

Tassew Tolcha

Abstract: Air transport, COVID-19, Africa, Aviation recovery COVID-19 is having multilayered impacts on African economies and the aviation industry. Countries are implementing various measures to help the industry survive the pandemic and facilitate resumption of air transport to pre-COVID-19 levels. However, air transport stakeholders still have concerns about the state of markets, the timing of recovery and recouping long-haul traffic. These concerns must be addressed in order to restore passengers' confidence in travelling, create air travel safety awareness and determine when to expect recovery. This study has two central aims: to estimate the immediate and long-term effects of COVID-19 on the air transport markets of nine African countries, and to forecast timescales for recovery of domestic and international flights. To realise these objectives, two related methods are applied – intervention analysis and SARIMAX – using monthly time-series data from August 2003 to December 2021. The empirical results show that air transport is significantly elastic to the pandemic. Lockdowns, travel restrictions and ensuing economic crises have resulted in reduced demand for air transport services. The combination of reduced travel demand, supply shocks and uncertainty about the long-term outlook create ambiguous prospects for airline companies. The analysis highlights that the immediate effects of the pandemic were higher in Morocco, Egypt, Mauritius and South Africa than in the other countries in the sample. Morocco, Egypt and Mauritius are the top tourist destinations in the region, and South Africa is the economic and air-transport hub for southern Africa, so air travel in these countries may be considered to be an essential mode of transport. However, the pandemic's long-term effects will be considerably more severe than its immediate effects, with significant differences across countries. Furthermore, more significant deviations are observed across countries in the timing of the effects (immediate versus long-term) than in types of markets (domestic versus international). This may relate to changing passenger behaviour following the COVID-19 crisis, the availability of alternative transport modes, trends in the pandemic and vaccination progress in each country. Estimates of recovery times show that air transport recovery will take on average 28 months (≈2.3 years) for domestic and 34 months (≈2.8 years) for international flights, starting from 2020. Tanzania's domestic flights are likely to recover quickest (13 months), followed by Kenya (22 months) and Ghana (24 months), while for international flights, Morocco is likely to have the shortest recovery period (≈2.3 years), followed by Egypt (≈2.6 years). This is in line with Gudmundsson et al.'s (2021) empirical estimate that world recovery of passenger demand to pre-COVID-19 levels will take 2.4 years. In general, the volatility of the air transport market following the outbreak of COVID-19 can be considered to be a cyclical business case rather than a structural change to the market. However, this does not mean that the market will be unaffected by behavioural changes by air transport passengers, accelerated by the transformation toward an e-society. COVID-19 has been episodic, so airlines are having to make rolling changes in response to unpredictable state measures. Moreover, further research is important to continuously monitor passengers' behavioural changes, industry responses and governmental policies, and to share relevant recommendations, as the COVID-19 crisis is not yet over. Future research might examine the pandemic's influence in accelerating the transformation towards an e-society, and its implications for post-pandemic air transport markets.

Rethinking Tourism and Hospitality Research

Mireya Morgana-Orellana, Patricia Picazo-Peral and Sergio Moreno-Gil

Abstract: Current literature has been conducting tourism research applying qualitative and quantitative approaches and methods to predict and forecast the future in the sector. After the COVID-19 pandemic, it is essential to analyze from a long term perspective where the tourism and hospitality research, as close related fields, are heading to. For this reason, future research in tourism and hospitality should consider quantitative and qualitative research from a holistic approach that integrates both tourism and hospitality, broadening horizons towards neurotourism research.

Using Forecasting Methods on Crime Data: The SKALA Approach of the State Office of Criminal Investigation North Rhine-Westphalia, Germany.

Kai Seidensticker and Katharina Schwarz

Abstract: 1 Introduction Crime is a complex social phenomenon which has a lasting impact on the population's sense of security and can burden society. The complexity of this phenomenon makes it therefore difficult to predict. Every day, new crimes of different types occur because of different decisions of perpetrators, resulting in many new trends in crime. To investigate these crime characteristics, crime patterns, and criminal behavior is a major objective of criminology. In the last decades, many new options in criminal investigation have aris-en as a result of rapid technical development and newly designed technologies. Because of the difficulty in dealing with the massive amount of data produced every day, data mining techniques have also become increasingly crucial in terms of criminal investigation. Nowadays, in the investigation, prosecution and prevention of crime, crime analysis plays a significant role. The intention of the SKALA (System for Crime Analysis and Anticipation) approach of the State Office of Criminal Investigation North Rhine-Westphalia (LKA NRW) is to understand the main use of crime analy-sis and forecast algorithms such as predictive policing methods, risk terrain modeling and time series anal-ysis, to investigate crime patterns at different spatial and temporal scales to support crime prevention. The initial findings are presented and discussed below. 2 Crime Forecasting in North Rhine-Westphalia In Germany, crime forecasting methods have been very virulent since 2014, especially connected to the term “predictive policing”. In the context of residential burglary, crime rates were increasing rapidly and led to the need of innovative approaches to tackle this crime phenomenon. Because in other countries predictive policing ought to be a successful method for reducing crime, this approach was also tested in Germany. Predictive policing still is an ill-defined subject (Seidensticker, Bode & Stoffel 2018). Available definitions range from calculating the probability of future crimes to specific methods to investigate crimes already committed (e.g. Perry et al. 2013). In Germany predictive policing is commonly defined as a computer-assisted method for spatially based probability calculations of crime (Seidensticker, Bode & Stoffel 2018). The aim is to identify risk areas so that the police can plan appropriate policing measures and optimally allocate their limited forces. Because Germany is divided into 16 states, each with its own state police force, some states have taken different approaches to predictive policing. In NRW, predictive policing was tested between 2015 and 2018 within the project SKALA to study the spatio-temporal dimen-sion of crime phenomena. One part of the project was to assess and predict the crime risk using spatial data for each residential district of the participating cities (LKA NRW 2018). After the successful completion of the project, SKALA was expanded into an independent research area within the Criminological Re-search Department of the LKA NRW and the crime forecasts were extended to the entire state of NRW in 2021. In SKALA, several options of (crime) forecasting methods are tested, such as a variant of the classic predictive policing approach for operational planning, risk terrain modeling and trend analysis for long-term predictions. 2.1 Predictive Policing Approach The predictions of the North Rhine-Westphalian variant of the classic predictive policing approach are based on a sequence of methods and processes. In a first step, data is selected, collected and prepared for further processing. Therefore, the police have information on past crime events in their databases that can be used for predictions. Crime events are mostly discrete events which are neither uniformly nor randomly distributed in space and time (Ratcliffe 2010). Their distribution is subject to different sociological theo-ries, such as the near repeat victimization approach (e.g. Polvi et al. 1991; Farrel & Pease 1993; Johnson et al. 2007) or the crime pattern theory (Brantingham & Brantingham 1995). In addition to crime data, socio-economic data is used to gain information about population structure, social status, but also on building structure and transport connections. In the next step on statistical procedures, such as modeling with sub-sequent forecasting, spatio-

temporal clusters were created from crime data using a score calculation utiliz-ing crime frequency, duration, and topicality which are aggregated over residential districts. After this, an automatic determination of important influencing variables was made using random forest model. Predic-tion and determination of space-specific influencing variables were performed using linear regression models. Following this, the predictions – which are computed every Monday and are valid for a timespan of one week – are processed and visualized for operational planning in the police authorities. 2.2 Risk Terrain Modeling Risk Terrain Modeling (RTM) tries to identify environmental characteristics that are linked to crime pat-terns (e.g. Perry et al. 2013). Because the RTM approach focusses the influence and distribution of socio-economic, building-specific, and infrastructure attributes and their distribution in space, this method might primarily be used by the police for regions with a very low number of cases in relation to the area (rural regions). In addition, the aspect of low numbers of crime events makes it difficult to use the classic predic-tive policing model. Statistical methods are used to identify risk factors from a large number of space-specific characteristics that significantly influence the risk burden (number of events per unit of space and time) for a selected offense. The mathematical model underlying the relationship between the significant characteristics and the risk burden subsequently enables a forecast calculation of the expected risk of crime. In this way, a risk map can be created for the entire space of a selected authority. The temporal va-lidity of such a map is linked to changes in the space-specific characteristics and the number of cases, which in rural regions are subject to rather minor changes on short time scales. In RTM, calculations are usually performed using the "unsupervised learning" method. This method is intended to identify previously unknown correlations or patterns without prior categorization of the data (e.g. Hastie, Tibshirani, & Friedman 2009). The spatial risk profiles generated in this way are derived from the respective geographic characteristics of the space, which can be identified as risk factors for crime through unsupervised learning. Studies suggest that the prediction accuracy of the models created in this way is better than that of classic hot-spot models (Wang & Brown 2012). However, the correlations found in such models do not necessarily imply a meaningful and substantively justifiable causality. Initial results for NRW show that a spatial risk assessment is also possible for more rural police departments. Due to the high computation times of the model for the large spaces and the desire for a fine-scale grid, a seasonal risk assessment for the police authorities is aimed at here. 2.3 Time series analysis Time series analysis enable the calculation of trends which are used for long-term prediction of crime events to facilitate strategic planning in police agencies. In this context, a decomposition of time series into their additive components is often performed because they can exhibit a variety of patterns. A de-composed crime time series consists of the original time series, and the three decomposed parts with the estimated trend component, seasonal component, and irregular component (Feng et al. 2019). For NRW, the estimated trend component revealed an increase of residential burglary until 2015, followed by a steady decrease until today. Regarding the seasonal component, an increase in the winter terms was de-tected and a decrease in the summer terms throughout the entire analysis period. From this, conclusions can be drawn for the following months as to the extent to which the case numbers of an offense are likely to increase, decrease or remain constant. A widely used time series forecasting model is Auto-Regressive Integrated Moving Average (ARIMA) introduced by Box and Jenkins (1970). The ARIMA model is a way to look at both long-term changes and seasonal components of crime. The use of the model is advantageous for crime data, as studies have shown that ARIMA works better compared to, e.g., complex Artificial New Networks in the case of a small num-ber of datasets (e.g., Jha et al. 2020). The fitted ARIMA model can be used to forecast future points in a series (Chen, Yuan, & Shu 2008). The experimental results produced for NRW using an ARIMA model of time series of residential burglary events support the value of crime time series analysis. Seasonal patterns were identified on a monthly, weekly, and even on a daily basis. The higher numbers of events in the win-ter months could also be mapped. 3 Potential and Limitations The use of spatio-temporally based statistical techniques offers police agencies the potential to optimize their operational and strategic planning based on the predictions to take necessary action before crime occurs. A targeted control of the forces or also the alignment of operational focal points on specific crime areas is made possible. Changes in crime can be detected earlier on the basis of such calculations. In evaluating crime forecasting methods, data quality is crucial concerning data uncertainty. Data uncer-tainty refers to the problem that it is usually unknown to what extent errors are contained in the collected and utilized data. In this context data collection problems, e.g. measurement uncertainties, are conceivable, e. g. when recording the suspected time of a burglary. For data collection in the context of police forces, a source of uncertainty is the fact that criminal offenses are either legally misjudged, or are reported late by the victims, which is not uncommon for the criminal charges of burglaries (Seidensticker, Bode & Stoffel 2018). Besides, the data quality of the data collected in the field of policing is often low, as time periods and localities are not (or cannot) always be precisely defined. Similarly,

it is difficult to predict offenses that should be prevented by increased police presence. Here, there is often a lack of knowledge as to whether police presence has prevented offenses or whether the predictive models are outputting inaccu-rate risk estimates. Overall, crime forecasting methods offer a wide range of possible applications that enable criminal exper-tise to be enriched by scientific findings.

Machine learning and satellite data processing for spatio-temporal forecasting - application to solar energy in Guyana

Sara Zermani, Jessica Bechet and Laurent Linguet

Abstract: Photovoltaic systems have many advantages such as the ability to be installed in remote locations. However, the efficient integration of PV production into the energy system is a major challenge due to the variability of solar irradiance. The solution to anticipate this variability is to provide accurate forecasts of the global solar irradiance on the ground. This will allow the planning, management and operation of photovoltaic installations in order to guarantee the stability and profitability of the power grid. In this paper, we propose an innovative spatio-temporal forecasting method from 1 to 12 hours ahead in the intertropical zone. Our work is based on data from the geostationary satellite GOES-16 that provides an image of the American continent every 10 to 15 minutes depending on the acquisition mode, with a spatial resolution of 0.5 km. Our objective is to reduce forecast errors, especially large errors due to the high variability of clouds in this area. The main idea is based on the selection of predictors extracted from satellite images. Once the predictors are extracted, we define the area of interest around the pixel based on statistical models. Then we build the machine learning model based on historical data (predictors, measured GHI, estimated GHI of the area of interest) of the pixel of interest. To validate our approach, we have used ground measurements and satellite-derived irradiance estimations from a station located in French Guyana and we have obtained a 3% improvement in RMSE.

The (a)symmetric aspects of business-consumer confidence on outward-inward tourism for Russia

Magdaline Enow Mbi Tarkang Mary

Abstract: Tourism activities in Russia (before 2019-2021, the coronavirus pandemic years) reportedly recorded a desirable boast, thus prompting the formulation of the country's new Tourism Development Strategy for 2035. Considering this observation, the current study examines the determinants of Russia's inbound and outbound tourism but especially from the perspective of the linear and nonlinearity role of business confidence and consumer confidence over the quarterly period 2010Q1-2020Q4. By applying linear and non-linear empirical analysis methods alongside Granger causality relationships from the more recent frequency domain causality approach, the investigation revealed interesting findings. The result shows that negative and positive shocks in business confidence and the consumer price index positively affect international arrivals to Russia in the short- and long-run, but a negative shock in business confidence shows a larger effect. A shock (irrespective of the direction) in the consumer leading indicator leads to a decline in international arrivals in the short- and long-run. We also found that a shock (mostly negative) in the consumer confidence reduces the number of Russian travelers to foreign tourist destinations in the short- and long-term periods. However, a negative shock in business confidence encourages Russians to embark on a touristic visit to destinations while there is no significant effect arising from the shock in the consumer price index. There was evidence of significant symmetric in the respective relationships as well evidence of Granger causality. For instance, there is a significant two-way Granger causality from BCI to outbound tourism, and from CLI to outbound tourism especially in the medium term while a one-way causality exists from CPI to outbound tourism in the short- and medium-term.

Exploration of Different Time Series Models for Soccer Athlete Performance Prediction

Siarhei Kulakou, Cise Midoglu, Nourhan Ragab, Matthias Boeker and Paal Halvorsen

Abstract: Professional sports achievements combine not only the individual physical abilities of athletes but also many modern technologies in areas such as medicine, equipment production, nutrition, and physical and

mental health monitoring. Proper diet, rest, and training regimens, as well as continuous monitoring and analysis of wellness and performance metrics can make a big difference for both individual and team sports. Soccer players constantly adhere to a strict nutrition plan, training process, and rest regime so that their body is in the required state at particular moments. The athletes' condition is influenced not only by the amount of consumed and burned calories, or the duration and intensity of the training process, but also by parameters such as the duration and quality of their sleep and their general mental state including mood and stress level. In this context, it is possible to compile a set of parameters describing the general state of the athlete at a certain point of time, collect objective or subjective measurements of these parameters, and try to predict the state/behavior of the athlete's body in the near future. In this work, we address the problem of predicting soccer players' ability to perform, from subjective self-reported wellness parameters collected using a commercially deployed digital health monitoring system called PmSys. We use 2 years of data from 2 Norwegian female soccer teams, where players have reported daily ratings for their readiness-to-play (0-10), mood, stress, general muscle soreness, fatigue, sleep quality (0-5), and sleep duration (hours). Using the Tsai library, we explore various time series models with the goal of predicting readiness, employing both a univariate approach, and a multivariate approach. We provide an experimental comparison of different time series models, such as purely recurrent models (RNN, LSTM, GRU, RNNPlus, LSTMPLus, GRUPlus), models of mixed recursive convolutional types (RNN-FCNPlus, LSTM-FCNPlus, GRU-FCNPlus), ensemble of deep CNN models (InceptionTime), and multivariate versions of the recurrent models (MRNN-FCNPlus, MLSTM-FCNPlus, MGRU-FCNPlus), in terms of prediction performance, with a focus on detecting peaks. We use windows of 7, 14, and 21 days as input, and prediction windows of 1 day and 7 days, allowing us to compare the accuracy of next-day predictions (which are less relevant in a professional soccer competition context, but interesting for exercise and training purposes) and next-week predictions (which are of interest to teams especially during the match season). We also investigate the potential of using models built on data from the whole team for making predictions about individual players, as compared to using models built on the data from the individual player only. As a common challenge associated with subjective daily report collection, we tackle the so-called "missing data problem" by various methods, including the replacement of all gaps with zeros, filling in average values, as well as removing all gaps and concatenating arrays. For evaluating the prediction performance, we use the metrics MSE, RMSE, and MAE. For evaluating the system performance, we consider disk space, memory, CPU, GPU, and time usage. Our case study on athlete monitoring shows that a number of time series analysis models are able to predict readiness peaks in most scenarios with a precision and recall of at least 90% in near real-time. Equipped with such insight, coaches and trainers can better plan individual and team training sessions, and perhaps avoid over training and injuries.

A Dynamic Combination of Optimized Theta Method with ATA, and Exponential Smoothing Methods; Validating on a Real Business Case and Comparing to Moving Average

Yasin Tadayonrad and Alassane Ballé Ndiaye

Abstract: Abstract. In order to make better decisions and take efficient actions in any supply chain system we need to have better estimation of the uncertain pa-rameters specially the future demand of our customers. To do so we must use a forecasting model which gives the most useful and accurate forecasts. Time series forecasting methods are still one of the most popular approach-es used in the business because of its simplicity. One of the most recent methods that caught the attention of researchers and practitioners is Theta method which was first ranked in M3 competition. This method works based on the decomposition of the deseasonalised original demand data in-to two components. The first component represents the long-term trend, and the second component indicates the short-term behavior of the data set. ATA method is another method has been introduced recently. ATA method works like Exponential Smoothing methods. But in ATA method the smoothing parameter is a function of time point. In this paper we have pro-posed a new form of Theta method in which we have benefited the features of ATA and presented a combination ATA method and Theta method. We have introduced a dynamic model which uses Theta method as the main model and select among some alternative methods such as ATA method, Simple Exponential, and Double Exponential smoothing methods to be used as the theta lines. Also, we optimize the parameters of each method used in the model. Finally, we have tested the mentioned model on a real data set and concluded that the combination of Theta and ATA method has a better performance compared to the other alternatives in terms of forecast accuracy.

Cloud Base Height estimation based on Convolutional Neural Network and All Sky Images

Emanuele Ogliari, Alfredo Nespoli, Dario Ronzio and Elena Collino

Abstract: Cloud Base Height (CBH) estimation is a crucial task: this parameter is required in many applications as it is exploited to validate climate models and to improve Numeric Weather Prediction (NWP) models. For instance, it improves the definition of the cloud-drift vector field, allowing a more effective modeling of the atmosphere dynamics. Moreover, CBH is also useful in the nowcasting (very-short term forecasting) of the solar radiation. As clouds are the primary cause of intermittency in solar irradiance, their characterization are extremely important for solar power applications: an accurate calculation of the ground shadowing require the knowledge of cloud height and extent. The objective of the present work is to estimate the CBH value starting from sky images acquired from a single ASI. In order to fulfill this task, a Convolutional Neural Network model is chosen and developed.

Alone we can do so little; together we can not be detected. Detection of outliers in time series by analyzing cluster transitions.

Sergej Korlakov, Gerhard Klassen, Stefan Conrad and Marcus Bravidor

Abstract: It is no longer possible to imagine our everyday life without time series data. This includes, for example, market developments, COVID-19 cases, electricity prices and other data from a wide variety of domains. An important task in the analysis of this data is the detection of anomalies. In most cases, this is done by looking at individual time series. In our work, we use the techniques of cluster analysis to establish a relationship between time series and groups of time series. This relationship allows us to observe the development of time series in their entirety and thereby gain additional insights. Our approach identifies outliers with a real-world reference and enables the user to locate outliers without prior knowledge. To underline the strengths of our approach, we compare our method with a known other method on two real-world datasets. We found that our solution needs a significantly lower number of calculations, produces more reasonable results, and can be applied to real-time data. Moreover, our method detected additional outliers, whose occurrence can be explained by real events.

Comparison of different portfolio optimization strategies to minimize the risk

Aleš Kresta and Anlan Wang

Abstract: In the scientific literature, there are various portfolio optimization models proposed with the aim to find the portfolio with the optimal risk-reward ratio: starting from Markowitz (1952) and his well-known mean-variance model, which laid the foundation of the portfolio optimization problem, continuing with the mean-CVaR models, until modern models such as the portfolio strategy maximizing Rachev ratio (Rachev et al., 2005). However, these optimization models depend on the forecast of the future price evolution, which means we need to estimate the parameters of the future returns' probability distribution. This kind of estimation may be subject to errors, see e.g., DeMiguel et al. (2009), and thus the observed out-of-sample performance usually differs from the in-sample historical performance. In this contribution, we compare the out-of-sample performances of various classical and modern portfolio optimization models on the basis of different risk and performance measures. When judging the portfolio performance, the commonly applied approach is to compare the performance of the portfolio with the benchmark. In our contribution, however, we base our comparison on the relative ranking of the portfolio performance among performances of uniformly distributed random portfolios. We take a similar approach as Kim and Lee (2016), who derived the equation for the relative ranking of the portfolio based on the Sharpe ratio and then tested the null hypothesis of whether the mean of the relative ranking is equal to 50%. In our contribution, due to the application of different risk and performance measures, we obtain the relative ranking via Monte Carlo simulation. In particular, we simulate the weights of portfolios such that these portfolios are uniformly distributed, then we calculate the risk and performance measures for each portfolio, and finally, we determine the relative ranking of our strategy portfolio among these random portfolios. Then we can test the null hypothesis that the mean relative ranking is equal to 50%. We use the

historical data of the Dow Jones Industrial Average index covering the period from January 3, 2006, to April 30, 2021. We also focus on the two special periods: the global financial crisis and the covid-19 pandemic, which allows us to study, how successful these models are when the market trend abruptly changes. We found out that the risk-minimization strategies generally decrease the risk in the out-of-sample period for all risk measures, i.e., the mean of the relative ranking is statistically lower than 50%. The strategies based on CVaR-minimization criteria decrease the risk and also increase the out-of-sample performance measures. We also found out that portfolio strategies maximizing the in-sample Rachev ratio do not decrease the out-of-sample risk, but increase the out-of-sample performance measures.

Recommendations of stockbrokers vs. fuzzy portfolio approach in construction sector

Aleksandra Wójcicka-Wójtowicz and Anna Łyczkowska-Hanćkowiak

Abstract: Short abstract. In portfolio management we can use oriented fuzzy numbers (OFNs) which include information uncertainty and imprecision related to financial market. In such a case we also utilize an expected fuzzy discount factor which leads to an imprecise present value (PV). Basing on those fac-tors we can obtain recommendations (buy, sell, accumulate, reduce) for in-dividual stocks included in the portfolio. Similarly, agents of stockbrokerage regularly issue analogous recommendation basing on their knowledge, experience and quantitative analysis. The main purpose of the paper is to analyse the published recommendations of stockbrokers, their accurateness and actu-al impact on the operations on portfolio including stocks of companies in construction sector. The created portfolio includes securities identified by a present value assessed with the use of trapezoidal oriented fuzzy number (TrOFNs) in times of issuance of the recommendation. All theoretical considerations are illustrated by an empirical case study. The elements of portfolio are positively or negatively oriented. Extended abstract 1 Introduction Individual investors before making an investment decision frequently search for reliable information on particular company or sector. Such information can take a form of recommendations of stockbrokers who are seen as professionals with broad knowledge and experience in this field. However, both knowledge and experience can be classified as imprecise and inaccurate foundations which, in turn, allows for the use of trapezoidal oriented fuzzy numbers in making portfolio investment decisions. In our paper, we investigate individual entities from construction sector included in a portfolio with PVs described by trapezoidal oriented fuzzy numbers. Furthermore, the analysis based on fuzzy discount factor is less difficult than the one based on a return rate. The paper is organized as follows. Section 2 presents trapezoidal fuzzy numbers, along with their main properties. In Section 3 the evaluation of imprecision for oriented fuzzy numbers is shown. Section 4 includes the notion of imprecisely estimat-ed PVs and expected discount factor. Section 5 compares the accurateness of bro-kers' recommendations and the recommendations resulting from the TrOFNs analy-sis. The research is illustrated by a case study. Section 6 concludes the findings of the research. 2 Methods Any security gives its owner a right to obtain future financial income that is payable at a specified maturity. The value of this gain can be presented as an anticipated future value of capital. However, as the present value of future cash flows is an ap-proximated value, it can be expressed by trapezoidal oriented fuzzy numbers which are simpler in case of operations. In our research a financial portfolio is understood as an arbitrary, finite set of assets. We consider both, a multi-asset portfolio and a single asset portfolio as an individu-al security can be assumed a portfolio. Any analysed asset is regarded as a fixed security in a long position. The considered portfolio elements are positively and negatively oriented. Their fuzzy discount factor is tested. Then, a weighted sum of positively oriented discount factors and a weighted sum of negatively oriented factors is calculated (due to the fact that adding oriented fuzzy numbers is not associative). Consequently, a portfolio discount factor is obtained as a weighted addition of both sums. In our procedure we use both entropy and energy measures for the estimation of the imprecision risk of obtained investment portfolios. 3 Findings The presented illustrative example shows that the brokers' recommendations in many cases were less accurate then the recommendations obtained from the fuzzy portfolio analysis for the companies in construction sector assessed by a forecast of the closest change in prices. We also observed that recommendations reached as a result of fuzzy portfolio analysis lower uncertainty and imprecision risk. 4 Conclusions The main objective of this paper was to compare the usefulness of the recommen-dations made by stockbrokers and suggested by the fuzzy portfolio analysis which was conducted with the use of fuzzy discount factor. The results show that the fuzzy portfolio analysis was more accurate for the companies in construction sector assessed by a forecast of the closest change in prices than the recommendations of the brokers. 5 References 1. Delgado, M.; Vila, M.

A.; Voxman, W. (1998) On a canonical representation of fuzzy numbers. Fuzzy Sets and Systems 93(1), 125–135. https://doi.org/10.1016/S0165-0114(96)00144-3 2. Dubois, D.; Prade, H. Operations on fuzzy numbers. Int. J. Syst. Sci. 1978, 9, 613–629. doi:10.1080/00207727808941724 3. Kosiński, W. (2006) On fuzzy number calculus. Int. J. Appl. Math. Comput. Sci., 16(1), 51–57 4. Łyczkowska-Hanćkowiak, A. On Application Oriented Fuzzy Numbers for Imprecise Investment Recommendations. Symmetry 2020, 12(10),1672. https://doi.org/10.3390/ sym12101672 5. Łyczkowska-Hanćkowiak, A. The Use of Trapezoidal Oriented Fuzzy Numbers in Port-folio Analysis. Symmetry 2021, 13, 1722. https://doi.org/10.3390/sym13091722 6. Łyczkowska-Hanćkowiak, A.; Piasecki, K. Present value of portfolio of assets with pre-sent values determined by trapezoidal ordered fuzzy numbers. Operations Research and Decision. 2018, 28, 41–56, doi:10.5277/ord180203 7. Łyczkowska-Hanćkowiak, A; Piasecki, K. Portfolio discount factor evaluated by orient-ed fuzzy numbers, 39th International Conference Mathematical Methods in Economics MME 2021. Conference Proceedings 8. Piasecki, K.; Łyczkowska-Hanćkowiak, A. Oriented Fuzzy Numbers vs. Fuzzy Num-bers. Mathematics 2021, 9, 523. https://doi.org/10.3390/math9050523 9. Piasecki, K. & Wójcicka-Wójtowicz, A. (2021). Application of the Oriented Fuzzy Numbers in Credit Risk Assessment. Mathematics 9, 5, 535. https://doi.org/10.3390/math9050535 10. Wójcicka, A. (2020). Influence of Imprecision on Credit Risk Assessment - Case Study. Proceedings of the International Scientific Conference Hradec Economic Days 2020, 871-880. 10.36689/uhk/hed/2020-01-098

An approximation to relative automatic time series forecasting measures

Miguel Ángel Ruiz Reina

Abstract: In this research, we discuss the accuracy of automatic models of the R programming package for univariate time series. The methods for selection criteria models used are proposed Lambda - Family - Measures to solve the problem of lack of definition widely mentioned in the literature. To solve this, we offer a reliable, robust and easy to interpret methodology for selecting models based on the Lambda theorem of accuracy for windows time horizon h= 3, 6, 12. The empirical results show no significant differences between predictive models for 53 series analysed from the tourism sector in Spain. Our results make it possible to empirically measure and quantify the accuracy of the models through multiple series and forecasting models. This proposed method-ology supposes a homogenisation for the forecasting model selection criteria comparisons.

High-Dimensional Temporal Disaggregation and Nowcasting: Examination of Trade-in-Services throughout the Brexit Transition Period

Luke Mosley, Alex Gibberd and Kaveh Salehzadeh Nobari

Abstract: In this work, we focus our attention on tracking imports and exports of quarterly UK Trade in Services throughout the Brexit transition period at a more granular (monthly) level than what is published. We develop a novel model that fuses together the high-dimensional temporal disaggregation method of Mosley et al. (2021) with the widely used two-stage nowcasting strategy of Doz, C., Giannone, D., & Reichlin, L. (2011). This fusion allows us to estimate a monthly disaggregate series of Trade in Services and nowcast monthly estimates forward up-until the next release of the output series. This way, we are able to look more closely into the short-term dynamics of the Services Trade and obtain early estimates of monthly movements before the next quarter is released. We then turn our attention into assessing the impact Brexit has had on the Services Trade. In this experimental work we try to disentangle the effects coming from the COVID-19 pandemic and from the Brexit transition using a difference-in-difference type approach based on our high-frequency nowcasts. The UKs Office for National Statistics (ONS) has made great strides in improving the timeliness of its key statistics, for instance, through integrating real-time-information and VAT data and the construction of a new monthly GDP index. However, many statistics of interest are not produced on a timescale relevant to understanding the fast moving economic environment that we have faced in the last few years. For instance, in the period of March 2020 - September 2021, the UK has moved into and out of nationwide lockdowns three times due to the COVID-19 pandemic. One key statistic that is of interest in this work is UK Trade in Services, a vital part of the UK economy, accounting for close to 80% of its GDP. The interest around this statistic has increased in the past few years mainly due to the UK voting to leave the EU and the subsequent restrictions of trade and movement of people. The UK officially left the EU on the 31st January 2020 and then entered a transition period. This period ended when the Trade and Cooperation Agreement (TCA) became active in January 2021. We are

interested in measuring Trade in Services before and after this transition. However, due to the 3-month publication delay, information on Quarter 1 of 2021 will only be available in June 2021. Such an information gap poses challenges for policy makers during these times of potential economic volatility. This lack of timeliness and delay suggests strong motivation for a methodology that is able to estimate statistics at a higher frequency than published and provide early high-frequency nowcasts to avoid long waits. This way, we endeavour to track the short-term dynamics of Trade in Services around the key month of January 2021. To produce monthly disaggregate estimates and real-time nowcasts of quarterly Trade in Services we use a variety of monthly indicator series coming from the Index of Production and Services, Trade in Goods and PPI and CPI data. In total, we have 675 indicator series that have all been seasonally adjusted. To use these indicators in our nowcasting exercise there are a few model considerations we must pay attention to. Our model has to be able to handle the mixed-frequency between the target and indicator series, it must also be able to handle the high-dimensionality and inevitable collinearity that exists between indicators. Further, indicators are released with a one or two month publication delay and thus we must handle this 'jagged edge' structure when producing nowcasts in real-time. To handle the above we fit a Dynamic Factor Model (DFM) onto the unbalanced indicator set to significantly reduce the high-dimension into a number of common factors that will be able to explain the majority of the variance in the indicators. The common factors are assumed to evolve over time following a VAR(p) model. We use the two-stage method of Doz, C., Giannone, D., & Reichlin, L. (2011) to estimate these factors. Preliminary estimates of the factor loadings and the factors are found by principle components analysis and a VAR(p) model is then fit to find autoregressive parameters. The model is then fully specified and cast into state space form where the Kalman Filter and Smoother is used to update the factors and estimate the missing observations where the 'jagged edge' is present. Then, using this balanced collection of monthly factors that are uncorrelated to each other and the quarterly Trade in Services target series, we fit the Sparse Temporal Disaggregation model of Mosley et al (2021) that is able to simultaneously produce a monthly disaggregate series of the quarterly target series over the observed quarterly sample and select the factors which are most significant in this disaggregation. This disaggregation method extends the well-established generalised least-squares approach of Chow and Lin (1971) by incorporating LASSO regularisation to allow for a high-dimensional indicator set and perform variable selection. Mosley et al (2021) shows the benefits this method has over Chow and Lin (1971) through several simulation studies. Using the estimated sparse LASSO regression parameter and the autoregressive residual parameter, we are able to forecast forward at the monthly level using our factor information estimated after the end of the quarterly sample. These forecasts are able to take into account the most up-to-date information observed as well as the autocorrelation present in the disaggregation residuals. It is hard to measure the accuracy of our monthly nowcasts since we do not observe the target at the monthly level. We perform an expanding window nowcasting exercise using data from Q1 1998 to Q3 2021 and each time aggregate the 3 out-of-sample monthly nowcasts that make up the quarter to compare to the observed figure. We maintain the 'jagged edge' structure of the indictors throughout the expanding window and re-estimate the whole model at each expansion to assess the variance in estimated factors. We also are able to assess the correlation of the monthly disaggregate estimates with monthly business survey data on trade in services. We find our model performs well against benchmarks. Current work is being done to zoom into the period around January 2021 and assess the impact Brexit has had on the imports and exports of Trade in Services. We attempt to disentangle the effects of the COVID-19 pandemic and the Brexit transition by using a difference-in-difference type approach. Quite a challenging task as both events occur within the same time window. We further look to incorporate higher-frequency real-time alternative data sources into the model to assist in the difficulty in nowcasting during 2020.

Estimating Growth at Risk with Skewed Stochastic Volatility Models

Elias Wolf

Abstract: This paper proposes a Skewed Stochastic Volatility (SSV) model to model time varying, asymmetric forecast distributions to estimate Growth at Risk as introduced in Adrian et al. (2019) seminal paper "Vulnerable Growth". In contrary to their semi-parametric approach, the SSV model enables researchers to capture the evolution of the densities parametrically to conduct statistical tests and compare different models. The SSV-model forms a non-linear, non-gaussian state space model that can be estimated using Particle Filtering and MCMC algorithms. To remedy drawbacks of standard Bootstrap Particle Filters, I modify the Tempered Particle Filter of Herbst and Schorheide (2019) to account for stochastic volatility and asymmetric measurement

densities. Estimating the model based on US data yields conditional forecast densities that closely resemble the findings by Adrian et al. (2019). Exploiting the advantages of the proposed model, I find that the estimated parameter values for the effect of financial conditions on the variance and skewness of the conditional distributions are statistically significant and in line with the intuition of the results found in the existing literature.

Performance Evaluation of Neural Network-Based Short-Term Solar Irradiation Forecasts

Simon Liebermann, Jung-Sup Um, YoungSeok Hwang and Stephan Schlüter

Abstract: Due to the globally increasing share of renewable energy sources like wind and solar power precise forecasts for weather data are becoming more and more important. When it comes to solar power, especially the quality of solar irradiation forecasts decides about the reliability of estimated future solar power amounts. Thereby, obviously, location matters. Forecasting models perform comparably better in areas with much sunshine and less clouds then in areas where clouds and/or precipitation are more likely (see Lorenz, 2009, or Mueller, 2010). Due to the above described demand there is already an extensive literature on forecasting solar irradiation and/or solar power, whereby lately the focus shifted towards methods that involve artificial neural networks (NN). In the context of solar irradiation forecasting we found more than 200 articles concerning with NN-based irradiation forecasts, whereby there are numerous suggestions to combine different NN setups with each other. Lately, various authors like Kreuzer (2020) or Wang (2019) started to combine a convolutional neural network (CNN) with a long-short term (LSTM) neural network in order to incorporate interdependencies in time and between different climate data like wind, temperature and irradiation. Kreuzer (2020), who applied a NN for generating short-term temperature forecasts, showed that this reduces forecasting errors compared to a pure CNN or LSTM network. On the one side this approach should be also beneficial for irradiation forecasting, as irradiation is more deterministic than temperature; for example, we have a clearly specified annual and daily pattern. On the other side, solar irradiation on the ground depends on highly stochastic factors such as cloud coverage, which is both an autoregressive and a highly volatile stochastic process. Wang (2019) applied such a combination to forecast 5 minutes irradiation values and showed that a combination of CNN and LSTM network indeed improves forecasting quality: In their article the method outperforms various NN benchmarks. However, their results have only limited validity, as they used one data set from a rather sunny location, namely Alice Springs in Australia, as a test set. Here, among other things, cloudiness is of minor importance. In order to see if the application of such a complex model really pays off, the method has to be tested in different climate conditions. Models in general perform worse in areas with less sunshine, because weather is less stable there. Lorenz (2009), for example, computed a forecasting error for Southern Europe between 20% and 35% whereas in central Europe the error increased up to 60%. Hence, a good performance in a sunny area is not necessarily a strong argument for a forecasting method. Up till now, when it comes to solar irradiation forecasting, there has been no efforts to (a) compare NN models of different complexity under different weather conditions and (b) analyze more in details the resulting error statistics. Authors commonly apply standard statistical measures like the root mean square error or various score measures. However, there is rarely an error analysis in the temporal domain. In this article we aim to fill this gap by including four different locations across Europe from sunny Almeria in Spain via the windy sea town of Hull to Rovaniemi in Finland, located close to the Arctic circle. Even though this still not covers all world-wide weather scenarios it give at least a good impression of how the models work under different conditions. It would be valuable information to know under which (weather) conditions a model performs good or not, and if the comparably high complexity of combining a LSTM and CNN model really pays off regarding forecasting precision. This is why we especially focus on evaluating the performance and analyze the errors also considering the time dimension. The setup of our case study is as follows: For all locations we produce hourly, i.e. short-term forecasts for global irradiation on the ground for up to 24h in advance. As auxiliary variables for the forecast we add various meteorological data such as temperature and rainfall. Results are analyzed in detail using standard goodness of fit measures. Errors are checked for bias and skewness. Above that we perform an analysis in the time domain to evaluate when and under which conditions errors are high or low. As expected, errors increase with the forecasting horizon. This makes sense as a longer horizons offers more chance for changing weather. Albeit the combination of LSTM and CNN performs better than other NN benchmarks on the long run, it is outperformed at all locations for short horizons. Besides, for Ulm, Hull and Rovaniemi, even for forecasting horizons larger than about 5h

the comparative advantage regarding precision does not justify the additional complexity. Interestingly, for Almeria, the sunniest candidate, combining CNN and LSTM network does make a change, as for the other locations, a comparably simple and well-established LSTM shows a reasonable performance. In Almeria it does not. Hence, the strength of the hybrid model shows in its versatility. We also see that all tested models are unbiased - except for the CNN, which on average overestimates irradiation a bit. Forecasting errors, which are computed as true irradiation minus forecasted irradiation are slightly skewed to the left, meaning that if we underestimate irradiation, the risk of significantly missing the real value is larger than when overestimating. Literature Lorenz, E.; Hurka, J.; Heinemann, D.; Beyer, H.. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2009, 2(1), 2-10. Müller, S.; Remund, J. Advances in radiation forecast based on regional weather models MM5 and WRF. Proceedings of the 25th European Photovoltaic Solar Energy Conference, Valencia, Spain, 2010. Kreuzer, D.; Munz, M.; Schlüter, S. Short-term temperature forecasts using a convolutional neural network — An application to different weather stations in Germany. Machine Learning with Applications, 2020, 2, 100007. Wang, K.; Qi, X.; Liu, H. Photovoltaic power forecasting based LSTM-Convolutional Network. Energy, 2019, 189.

Effects of Covid on the performance of different prediction methods: fare revenue forecasting in public transportation in Berlin

Nicki Lena Kämpf, Jonas Krembsler, Sandra Spiegelberg, Nicola Winter, Thomas Winter and Robert Knappe

Abstract: We present first results from a case study of fare revenue forecast during Covid in public transportation in Berlin. The data is based on monthly fare revenues for different product segments. The results will be used in a research project in public transport with the goal of automating revenue controlling and implementing data-driven decision-making in the existing controlling processes. The focus of this study is to examine how the performance of prediction methods changes and how long different methods need to adjust to the Covid shock. Therefore, the fare revenue predictions in public transportation for the years 2018, 2019, 2020 and 2021 are compared based on different accuracy measures. The study covers a wide range of methods such as SARIMA(X), Holt-Winters, Multiple Linear Regression, Random Forest, Gradient Boosting and Neural Networks.

Performance evaluation of a family of GARCH processes based on Value-at-risk forecasts

Alex Babiš and Beáta Stehlíková

Abstract: Financial time series forecasting focuses mainly on predicting the movement of underlying assets, such as stocks, bonds, commodities, etc. However, in specific cases, the prediction of the movement is not of interest, but instead the focus is on forecasting the future distribution of assets. Distribution forecasting is probably most commonly applied in the field of financial risk management where the future losses of investments are quantified by risk measures such as Value-at-risk. Both underestimation and overestimation of Value-at-risk are issues from an economic point of view and forecasting accurate Value-at-risk measure is desired. Our interest lies in the predictive power of some Value-at-risk models evaluated via Backtesting procedures that quantifies the adequacy of methods based on historical data (Kupiec test, Christoffersen test, Dynamic quantile test, Loss functions etc.) and their complex comparison. Financial time series often exhibits volatility clustering, i.e. volatility changes over time and period with low or high volatility can be observed. Common models used to capture volatility dynamics belong to a family of discrete GARCH processes. Apart from basic GARCH process, we consider two processes that are able to capture persistence of volatility (integrated GARCH and fractionally integrated GARCH process), three processes that can incorporate asymmetric information of positive and negative shocks (exponential GARCH, Glosten-Jagannathan-Runkle GARCH and threshold GARCH process) and one process that allows us to model short- and long-run movements of volatility (component GARCH process). Beside standard normal distribution, we also consider three additional distributions of innovations, which are helpful to incorporate known phenomena associated with financial time series modeling. Student t distribution is used to model heavy tails, and skewed normal distribution is used to model asymmetry of distribution of returns. Finally, skewed student t distribution was

chosen to capture both characteristics at the same time. Final models are combinations of one of the assumed distributions with one of the assumed volatility specifications. The aim is to find a subset of models which are better than the rest. The estimation of parameters and Value-at-risk forecasts are carried out on constituents of EURO STOXX 50 index. It is reasonable to expect that we would not find a subset of models performing better than the rest considering all criteria and time series, simply because of diversity and quantity of the data. An appropriate partial aggregation of the results for each criterion can lead to a simplification of the evaluation. The aggregated criteria can be then passed as inputs to data envelopment analysis models that serve as a framework to assess efficiency and ultimately lead to a set of – in a sense – preferred models.

On the correlation analysis of illiquid stocks

Hamdi Raissi

Abstract: The serial correlations of illiquid stock's price changes are studied, allowing for unconditional heteroscedasticity and time-varying zero returns probability. Depending on the set up, we investigate how the usual autocorrelations can be accommodated, to deliver an accurate representation of the price changes serial correlations. We shed some light on the properties of the different serial correlations measures, by mean of Monte Carlo experiments. The theoretical arguments are illustrated considering shares from the Chilean stock market.

Filling the Gap between Continuous Time AR Processes and Discrete Time AR Sequences. Application to Time Change Issues.

Rachid Senoussi and Valérie Girardin

Abstract: In theory, continuous time processes may be observed through trajectories on intervals. In practice, data depending on time are mainly collected – regularly or irregularly, sparsely or densely – at discrete observation times, and hence chronologically ordered by integer numbers, finally yielding discrete time sequences. The other way round, many discrete time sequences can be considered as originating from a continuous time process with data sampled at some pertinent time scale. Figuring out how the discretely sampled process inherits the properties of the original continuous version is a classical issue. Conversely, figuring out how properties of a discrete sequence can be used to construct continuous versions yielding back these properties to the ensuing discretized sequence is less classical but of equal interest. Both are detailed in [1]. In one way, observing the continuous time AR process at discrete times leads the AR dynamics of the discretized process to be characterized. The other way round, AR sequences are embedded, in the almost sure sense, into continuous time AR processes with the same dynamics. The communication will focus on filling this aller-retour between continuous time AR processes and discrete time AR sequences. AR models are here understood in a large sense, including time-varying coefficients, and semi-martingale type innovation, compound Poisson noise, in the continuous time case or non-Gaussian noise in the discrete case. Still, some hints will be given on associated time change issues. A scalar discrete or continuous time process is reducible to stationarity (RWS) if its transform by some smooth time change is weakly stationary. Different issues linked to this notion are investigated for AR models in [2] and [3]. In [2], any smooth time transform of a scalar continuous time AR process of ordre p, say CAR(p), with time dependent coefficients is shown to be a CAR(p) process too, with explicit underlying stochastic differential equation. Necessary and sufficient conditions are given for a CAR(p) to be stationary or to become stationary after undergoing a pertinent time change. For that purpose, a new criterion for weak stationarity is given, proven to be equivalent with the classical notion of asymptotic stability. In [3], following another point of view, necessary and sufficient conditions for scalar AR models to be RWS are investigated, with explicit formulas for the time changes. Stationarity reduction issues for discrete sequences sampled from time continuous AR processes are also considered. Several types of time changes and RWS processes are studied with examples and simulation, including the now classical multiplicative stationarity -- with logarithmic time change. References : [1] Filling the Gap between Continuous and Discrete Time Dynamics of Autoregressive Processes V. Girardin, R. Senoussi, Journal of Time Series Analysis V41, pp590--602 (2020) [2] Time Changes and Stationarity Issues for Continuous Time Scalar Autoregressive Processes V. Girardin, R. Senoussi, in Statistical Topics and Stochastic Models for Dependent Data with Applications, Ed. V. Barbu and N. Vergne, ISTE Wiley (2021) [3] Time Change for Scalar Autoregressive Models V. Girardin, R. Senoussi Work in Progress (2022)

Bridging fundamental and statistical models for short term electricity forecasting for the German market.

Souhir Ben Amor, Thomas Möbius and Felix Müsgens

Abstract: Electricity price forecasting is crucial for producers and consumers to plan their short-term bidding strategies and manage the risk related to the electricity market. However, due to the complex features of electricity prices, the forecasting task becomes even more difficult. Indeed, electricity prices usually present some form of non-stationarity, seasonality, and long memory behavior and they also depend on some fundamental factors such as weather, demand, and generation technologies. Hence, a robust model that takes into consideration all the mentioned features is needed for accurate electricity price forecasting. In this paper, a fundamental approach that includes techno-economic information with regard to generation and storage technologies, the demand side, and carbon emission allowances, is adopted to estimate the electricity prices by optimizing the production cost. However, the fundamental models are not able to provide short-term forecasting due to their weakness in capturing the short-term price dynamics. To overcome this limitation, we propose to build a bridge between the fundamental approach and an econometric model. The econometric approaches can take into consideration the statistical characteristics related to the electricity market (such as high volatility, long memory behavior, non-linearity, etc). Moreover, they are widely used for short-term forecasting. Despite their efficiency, the econometric models suffer from some drawbacks. First, they assume that the historical price includes all information needed to forecast the future, in another world, they assume that "the history repeats itself". Second, they are not able to incorporate the market dynamics and market operations into their price forecasts. Therefore, combining the fundamental approach with an econometric model allows us to take into consideration the most relevant economic factors, and also several statistical aspects when forecasting the electricity prices. For this purpose, the estimated electricity prices from the fundamental model are used as an input (with other explanatory variables such as wind forecasting,..) for the Lasso Estimated AutoRegressive model to predict the short-term electricity price in the German market. The predictive accuracy of the proposed model is then compared with the individual components separately and with the Deep Neural network and Random forest models. Finally, we determine what value the fundamental electricity market model has for price forecasts in such a bridging approach.

PV fault diagnosis method based on time series electrical signal analysis

Carole Lebreton, Fabrice Kbidi, Frédéric Alicalapa, Michel Benne and Cédric Damour

Abstract: With the objectives of energy self-sufficiency and zero emissions in La Reunion, photovoltaics are becoming an increasingly important part of the local energy mix. Installations reliability and safety are crucial to ensure network stability and security, therefore PV system fault diagnosis are essential tools in the expansion of this electricity production mean. DETECT Project (Diagnosis onlinE of sTate of health of EleCTric systems) is a research project aiming at diagnosis methods development. In this way, signal processing provides us promising tools as decomposition algorithms. Thanks to its low computation cost, Variational Mode Decomposition (VMD) allows undertaking a real-time diagnosis, with on-line PV electrical signals time-series data analysis.

An Example of Time-Series Modelling with Bayesian Symbolic Regression

Krzysztof Drachal

Abstract: Bayesian Symbolic Regression (BSR) is a novel time-series modelling method. The already known, symbolic regression with genetic programming (SR-GP) is a time-series modelling method basing on certain algorithms. Suppose that there exists a certain initial set of functions. Let these functions be, for example, multilinear regression models, that can be constructed out of the given set of explanatory variables. They can be, for example, all possible linear models, some subset of them, or some models with non-linear features. In other words, if there is a set of explanatory variables and the set of operators (like, for linear models: addition, and for non-linear ones: additionally some others, like, multiplication, logarithms, etc.), then a set of functions can be constructed. SR-GP uses genetic programming to evolve the initial set of functions with a hope to find

some “optimal” solution. For this purpose, it uses mutation and cross-over procedures, as well as, the new population of functions is somehow reduced to limit the size. In each step the functions are evaluated due to some criterion, for example, Root Mean Square Error (RMSE). After some steps, it is hoped that the remaining set contains the “optimal” solution. Recently, BSR has been proposed, which is a Bayesian method. Instead, of genetic programming, this method explores the populations of functions and seeks the “optimal” solution by using Bayesian regression trees models. Technically, it applies the Metropolis-Hastings algorithm and Markov chain Monte Carlo (MCMC) methods. From the practical point of view, this method can be applied to the variable selection problem in forecasting crude oil spot price. As an example monthly data between 1989 and 2021 have been analyzed. WTI spot price has been taken as the response variable. Although, for robustness checks also the Brent and the Dubai prices have been used. The possible explanatory variables have been taken as world production of crude oil including lead condensate, OECD refined petroleum products consumption, U.S. ending stocks excluding Strategic Petroleum Reserves of crude oil and petroleum products, MSCI World stock market index for developed markets, VXO index, the Kilian index of global economic activity and real narrow effective exchange rate for U.S. Additionally, Chinese stock market index has been taken. This index has been constructed by the suitable gluing of Hang Seng index (before 1991) and SSE Composite index (after 1991). In BSR the “optimal” solution is assumed to be the sum of a certain number of components. The components themselves can be non-linear. Herein, between 1 and 9 components have been tested, as this number must be preselected for initialization of BSR algorithm. Secondly, 7 different functions sets have been tested. The obtained forecasts from numerous BSR models have been compared by the Model Confidence Set test (MCS) in case of multiple forecasts comparison or by the Diebold-Mariano test for pairwise comparison of forecasts accuracies. The forecast accuracy has been measured by RMSE. Besides, various other modelling methods able to deal with variable selection have also been performed. In particular, Dynamic Model Averaging, Bayesian Model Averaging, Dynamic Model Selection, Bayesian Model Selection, LASSO and ridge regressions (standard and Bayesian versions), elastic net and least-angle regression. Moreover, time-varying parameters regressions, ARIMA, historical average and the naïve (no-change) forecasts have been estimated. BSR has produced significantly more accurate forecasts than SR-GP in most of the cases. Moreover, BSR has been found to be significantly more accurate than the naïve method. On the other hand, the clear pattern for specification of initial parameters of BSR (number of components and set of functions) has not been found. According to MCS, BSR has remained amongst superior (in a sense of forecast accuracy) models for all three crude oil spot prices. The other superior models have been ARIMA and Dynamic Model Averaging, depending on which price (WTI, Brent or Dubai) has been considered. Acknowledgements Research funded by the grant of the National Science Centre, Poland, under the contract number DEC-2018/31/B/HS4/02021. Literature CBOE, 2021, VIX Historical Price Data, https://www.cboe.com/tradable_products/vix/vix_historical_data Diebold, F. X., Mariano, R. S., 1995, Comparing predictive accuracy, Journal of Business and Economic Statistics 13, 253-263. EIA, 2021, U.S. Energy Information Administration, https://www.eia.gov FRED, 2021, Economic data, https://fred.stlouisfed.org Friedman, J., Hastie, T., Tibshirani, R., 2010, Regularization paths for generalized linear models via coordinate descent, Journal of Statistical Software 33, 1-22. Hansen, P. R., Lunde, A., Nason, J. M., 2011, The Model Confidence Set, Econometrica 79, 453-497. Hyndman, R. J., Khandakar, Y., 2008, Automatic time series forecasting: the forecast package for R, Journal of Statistical Software 26, 1-22. Jin, Y., Fu, W., Kang, J., Guo, J., Guo, J., 2019, Bayesian symbolic regression, https://arxiv.org/abs/1910.08892 Kilian, L., 2009, Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, American Economic Review 99, 1053-1069. Koza, J. R., 1992, Genetic Programming, Cambridge, MA: MIT Press. MSCI, 2021, End of Day Index Data Search, https://www.msci.com/end-of-day-data-search Nicolau, M., Agapitos, A., 2021, Choosing function sets with better generalisation performance for symbolic regression models, Genetic Programming and Evolvable Machines 22, 73-100. Raftery, A. E., Karny, M., Ettler, P., 2010, Online prediction under model uncertainty via Dynamic Model Averaging: Application to a cold rolling mill, Technometrics 52, 52-66. Stooq, 2021, Quotes, https://stooq.com The World Bank, 2021, Commodities Markets, https://www.worldbank.org/en/research/commodity-markets

Expectation-maximization algorithm for autoregressive models with Cauchy innovations

Monika Singh Dhull and Arun Kumar

Abstract: In this paper, we study the autoregressive (AR) models with Cauchy distributed innovations. In the AR models the response variable y_t depends on previous terms and a stochastic term (the innovation). In the classical version, the AR models are based on normal distribution which could not capture the extreme values or asymmetric behavior of data. In this work, we consider the AR model with Cauchy innovations which is heavy-tailed distribution. We derive closed forms for the estimates of parameters of the considered model using expectation-maximization (EM) algorithm. The mean and variance of Cauchy distribution are not defined, therefore we analyze the behavior of the time series using the characteristic function of Cauchy distribution. The efficacy of the estimation procedure is shown on the simulated data. The comparison of the proposed EM algorithm is shown with the maximum likelihood estimation method.

Towards time series feature engineering in automated machine learning for multi-step forecasting

Can Wang, Mitra Baratchi, Thomas Thomas Bäck, Holger H. Hoos, Steffen Limmer and Markus Olhofer

Abstract: Feature engineering is an essential step in the pipelines used for many machine learning tasks, including time-series forecasting. Although existing AutoML approaches partly automate feature engineering, they do not support specialised approaches for time-series data such as multi-step forecasting. Multi-step forecasting is the task of predicting a sequence of values in a time series. Two kinds of approaches are commonly used for multi-step forecasting. A typical approach is to apply one model to predict the value for the next time step, then the model uses this predicted value as an input to forecast the value for the next time step. Another approach is to use multi-output models to make the predictions for multiple time steps of each time-series directly. In this work, we demonstrate how automated machine learning can be enhanced with feature engineering techniques for multi-step time-series forecasting. Specifically, we combine a state-of-the-art automated machine learning system, auto-sklearn, with tsfresh, a library for feature extraction from time-series. In addition to optimising machine learning pipelines, we propose to optimise the size of the window over which time-series data are used for predicting future time-steps. This is an essential hyperparameter in time-series forecasting. We propose and compare (i) auto-sklearn with automated window size selection and (ii) auto-sklearn with tsfresh features. We evaluate these approaches with statistical techniques, machine learning techniques and state-of-the-art automated machine learning techniques, on a diverse set of benchmarks for multi-step time-series forecasting, covering 20 synthetic and real-world problems. Our empirical results indicate a significant potential for improving the accuracy of multi-step time-series forecasting by using automated machine learning in combination with automatically optimised feature extraction techniques. Our methods yield improved accuracy for 18 out of 20 datasets compared with the state-of-the-art.

Improving the Predictive Power of Historical Consistent Neural Network

Rockefeller Rockefeller, Bubacarr Bah, Vukosi Marivate and Hans-Georg Zimmermann

Abstract: The Historical Consistent Neural Network (HCNN) is an extension of the standard RNN: they allow the modeling of highly- interacting dynamical systems across multiple time scales. HCNN does not draw any distinction between inputs and outputs, but model observ- ables embedded in the dynamics of a large state space. In this paper, we propose to show that HCNN improve the predictive power of HCNN using three methods:(1) Partial Teacher Forcing, (2) Large Sparse State Transition Matrices, and (3) a Long Short Term Memory Formulation of HCNN. We investigated the effect of those long memory improve- ment methods on time-series datasets mathematically generated from the Lorenz system. To complement our study, we also compared the accuracy of different HCNN variants with well-known recurrent neural networks methods such as Vanilla RNN and LSTM for the same predic- tion tasks. Our results obtained show that the Vanilla HCNN has better results than RNN and LSTM when presented with the forecast of such dynamical system. Moreover, the forecast accuracy of the Vanilla HCNN has been improved by a factor of 3 by the improvement methods, as pre- sented in our findings. We demonstrate that (1) and (3) are superior for the modeling of our chaotic

dynamical systems. To show that the results would not differ too much for different instances of HCNN, we show the results in form of ensemble computations.

Multi-Country and Multi-Horizon GDP Forecasting Using Temporal Fusion Transformers

Juan Laborda, Sonia Ruano and Ignacio Zamanillo

Abstract: This paper applies a new artificial intelligence architecture, the Temporal Fusion Transformer (TFT), for the joint GDP forecasting of 25 OECD countries at different time horizons. This new attention-based architecture offers significant comparative advantages over other deep learning methods. First, results are interpretable since the impact of each explanatory variable on each forecast can be calculated. Second, it allows to visualize persistent temporal patterns and to identify significant events and different regimes. Third, it provides quantile regressions and permits to train the model on multiple time series from different distributions. Results suggest that TFTs outperform regression models, especially in periods of turbulence, such as the Great Recession or the Covid-19 shock. In essence, TFT is revealed as a new tool that artificial intelligence provides to economists and policy makers, with enormous prospects for the future.

Deep Representation Learning for Cluster Level Time Series Forecasting

Tsegamlak Terefe Debella, Bethelhem Seifu Shawel, Maxime Devanne, Jonathan Weber, Dereje Hailemariam Woldegebriel, Sofie Pollin and Germain Forestier

Abstract: In today's data-driven world, time series forecasting is one of the dominant data mining techniques associated with temporal datasets. In practice, there is a range of forecasting techniques that are proven to be efficient at capturing different aspects of their inputs. For instance, the classic linear Seasonal Auto-Regressive Integrated Moving Average (S-ARIMA) model is known to capture trends and seasonality evident in data. Moreover, neural network-based approaches are known to be best at capturing the correlation among time stamps. In practice, most forecasting techniques inherently assume that models are fitted using a single input. However, in reality, there are often cases where we cannot deploy forecasting models on individual series for various reasons. For instance, in most wireless communication traffic forecasting problems, temporal datasets are defined by taking samples from hundreds of base stations that have spatial correlations due to user mobility, land use, settlement patterns, etc. Thus, in such cases, it is often advised to approach forecasting using clusters that group base stations based on their traffic pattern prior to prediction. However, when this is the case, the quality of the cluster centroids and the overall cluster formation process is expected to have a significant impact on the overall performance of forecasting models. In this paper, we show the effectiveness of representation learning on cluster formation and cluster centroid definition which in turn improves the quality of cluster-level forecasting. We demonstrate this concept using data traffics collected from 729 wireless base stations. In general, based on the experimental results, the representation learning approach outperforms cluster level forecasting models that were based on classical clustering techniques such as K-means and Dynamic Time Warping Barycenter Averaging K-means (DBA K-means).

On Generating Synthetic Time Series For Cybersecurity

Veronica Sirghei

Abstract: Real-world data in many domains is difficult to obtain for privacy reasons. One solution for the lack of public datasets is to perform experiments using synthetic data. In this paper, we propose a comparison between Generative Adversarial Network based approaches for generating synthetic time series with practical applications in the domain of cybersecurity. We conduct experiments using several cybersecurity datasets to demonstrate that relevant synthetic datasets can be produced with high utility and strong similarity to the real datasets.

Time Series Forecasting Using Smoothing Ensemble Empirical Mode Decomposition and Machine Learning Techniques

Mona Bahri, Sahand Vahidnia and Mahtab Mohtashamkhani

Abstract: Forecasting is one of the most important features in time series analysis. When the time series is non-stationary and non-linear, forecasting is even more challenging. Recently, machine learning techniques are widely used to predict time series. However, for non-stationary and non-linear data, machine learning alone fails to predict the correct features of the time series. In this paper, to improve the results of forecasting for non-stationary and non-linear signals a combined prediction method is suggested. This method is a combination of the smoothing ensemble empirical mode decomposition (SEEMD) and machine learning techniques (the long short-term memory (LSTM) networks and convolutional neural networks (CNN)). A range of synthetic non-stationary and non-linear data, including data with no trend, linear trend and non-monotonic trend, and noise, are used to evaluate the accuracy of the predicted results. The advantage of using synthetic data is that the actual signal components are known, which permits a direct comparison of the predicted signals with the actual data. 70% of this synthetic data is used for the experiment and the training steps of the machine learning process and the last 30% is reserved to be compared with the predicted time series. In this paper, first, the SEEMD method is used to decompose the data into its embedded modes, called intrinsic mode functions (IMF). This will eliminate the possible noise and outliers in the data. Then, the machine learning technique is used for each IMFs. In the experiment phase of machine learning, several methods are used on each IMFs and their prediction results are compared with the original IMFs to select the most suitable prediction method for each IMFs. The experiment showed that the LSTM network model is the most suitable model for predicting sinusoidal (high and low frequency) and CNNs prediction is the most suitable for predicting trends. Then, a combination of LSTM networks and CNNs models then established for all the IMFs to train and forecast the time series. Finally, the prediction results for each IMFs are extracted and compared with the original synthetic data. The following table shows the R2 and MAE for all IMFs and all the data: The above table shows, that the proposed method of SEEMD and machine learning produce accurate results (with R2 above 0.9) for all the IMFs and all the data that is been used in this study. For all synthetic data, the predicted low frequencies were very accurate with MAE of 0.0018 to 0.0023 and R2 = 0.9999. However, the predicted high frequency IMFs had a larger errors (MAE= 0.04). This is due to the SEEMD inabilities to fully separate the high-frequency modes from noise, which renders the generalization capability of the models to be less effective. This error becomes apparent by over and under-estimations around local extremum points. For the trend, although the R2 is quite high, there was some evidence that added non-linearity increased the trend error (by MAE= 0.00154), especially at the endpoints. This might seems a minor error but indicates that further study is needed for data with a sharper trend. In summary, the proposed method is a suitable method to forecast nonstationary and non-linear time series. The finding in this paper shows that the LSTM network is suitable for predicting high and low frequencies as it can learn and predict signals with sinusoidal patterns. High-frequency Low-frequency Trend Data MAE R2 MAE R2 MAE R2 Linear Trend 0.0434 0.92183 0.00226 0.99995 0.0002 0.99999 Non-mono Trend 0.04254 0.92511 0.00167 0.99997 0.00174 0.99998 No Trend 0.04328 0.9223 0.00179 0.99997 - -

Heterogeneous agent model of multi-assets in discrete-time

Serena Brianzoni, Giovanni Campisi, Graziella Pacelli and Filippo Petroni

Abstract: We start from the continuous-time heterogeneous agent model of two risky assets of [6]. Our aim consists in studying the economic time-varying dominance in financial market and the other stylized facts, in discrete time. To this end, the continuous-time deterministic skeleton of the model [6] will be transformed using sophisticated discretization techniques. After that, we carry out the stability analysis of the resulting nonlinear dynamical system and focus on the local and global bifurcations leading to the transition to complex dynamics (see, among others, [4] and [5]). We will see as heterogeneity introduces an important nonlinearity into the market dynamics, causing our deterministic financial model to generate erratic time series similar to the patterns observed in real markets. In particular, following [7] and [8], we demonstrate that the originated time series are in line with the S&P 500 in terms of non-normality of returns, volatility clustering and long range dependence on volatility (see e.g. [1], [2] and [3]). Additionally, a comparison with the benchmark model of [6] will be made, in order to stress similarities and new evidences. References [1] Brianzoni, S., Cerqueti, R., Michetti, E.: Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Computational Economics, 35,

165-188 (2010) [2] Brianzoni, S., Mammana, C., Michetti, E.: Updating Wealth in an Asset Pricing Model with Heterogeneous Agents. Discrete Dynamics in Nature and Society, 2010 (2010) [3] Brianzoni, S., Campisi, G.: Dynamical analysis of a financial market with fundamentalists, chartists, and imitators. Chaos, Solitons & Fractals, 130, 109434 (2020) [4] Brock, W., Hommes, C.: Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control, 22(8-9), 1235-1274 (1998) [5] Chiarella, C., He, X.Z., Huang, W., Zheng, H.: Estimating behavioural heterogeneity under regime switching. Journal of Economic Behavior & Organization, 83(3), 446-460 (2012) [6] He, X.Z., Li, K., Wang, C.: Time-varying economic dominance in financial markets: A bistable dynamics approach. Chaos: An Interdisciplinary Journal of Nonlinear Science, 28(5), 055903 (2018) [7] He, X.Z., Li, Y.: Power-law behaviour, heterogeneity, and trend chasing. Journal of Economic Dynamics and Control, 31(10), 3396-3426 (2007) [8] Lux, T., Alfarano, S.: Financial power laws: Empirical evidence, models, and mechanisms. Chaos, Solitons & Fractals, 88, 3-18 (2016)

An ordinal procedure to detect change points in the dependence structure between non-stationary time series

Alexander Schnurr and Svenja Fischer

Abstract: We present a procedure to detect changes in the strength of dependence between data sets. Our method is based solely on a comparison of the ordinal structures within a moving window and thus compares the up-and-down behavior only. Hence, it is not distracted by changes within a single data set, such as change-points and trends or even non-linear transformations, leading to non-stationarity. The applicability of the method is demonstrated for a hydrological data set of runoff time series which are impacted by a reservoir.

Modelling a Continuous Time Series from FOU(p) processes

Juan Kalemkerian

Abstract: In this work we present the fractional iterated Ornstein–Uhllenbeck Processes of order p (FOU(p)). We give its definition, the main theoretical properties and a procedure to estimate consistently its parameters. We also show how to model a continuous time series from FOU(p) processes.

Nowcasting US-Inflation with random forests

Wegmüller Philipp

Abstract: We develop a random forest model which allows for mixed frequencies and missing observations. Further, we generalize the expected value of the target variable at each node of the regression tree to allow for a linear relationship between target and predictors. We apply the proposed framework to nowcast US inflation using a large set of daily, weekly and monthly data. The resulting predictions are significantly better than simple univariate benchmarks for US inflation rate indices and also outperform the nowcasts published by the Cleveland Fed.

Three states of the French business cycle

Catherine Doz and Anna Petronevich

Abstract: In this paper, the French business cycle is modelled through a Markov Switching Dynamic Factor Model. A one-factor model is estimated on a large database of macroeconomic series, using a two-step approach. The factor is modelled as a Markov-Switching process, for which several specifications are considered. It appears that the best specification is a Markov-Switching process with three states - expansion, mild recession and severe recession, with severe recessions characterized as sharp but brief deterioration of economic situation. The resulting smoothed probabilities provide a correct dating of past French business cycles and detect well the waves of the Covid-19 crisis. The model can be used to nowcast the current state of the business cycle, and can be also useful to nowcast French GDP. However it improves only marginally the out-of-sample forecasts.

Short-Term Electricity Load Forecasting: Modelling the COVID-19 Lockdown in Spain

Eduardo Caro Huertas and Jesús Juan Ruiz

Abstract: Many governments around the world had to enforce a strict lockdown dur-ing the coronavirus COVID-19 pandemic, closing or limiting all non-essential businesses and ordering citizens to stay at home. As a result of this decision, the electric power demand pattern varied significantly during the affected months. Considering that the electric consumption forecasting models are estimated or trained using historical data, their numerical accura-cy decreased during months of confinement. In this work, the electric power consumption in Spain during the lockdown has been analyzed, and the forecasting software (currently used by the Spanish Transmission System Operator) has been modified to: (i) properly model the lockdown policies during the year 2020, and (ii) to anticipate the change of behavior of load consumption during working and non-working days. In order to evaluate the methodology's performance, a real data set from the Spanish market has been used and the developed models has been tested employing the RMSPE (Root Mean Squared Percentage Error) as the accuracy metric.

Testing for and measuring serial dependence by neural networks

Jinu Lee

Abstract: Testing serial dependence is central to much of time series econometrics. This paper focuses on the generalisation of an autocorrelation function to test for and measure serially dependent processes by using neural networks based approximations. Simulations find that the suggested nonparametric method shows good power properties and has the potential to measure nonlinear associations compared to some popular tests and measures. An application to US stock returns illustrates the usefulness of the proposed tests and measures for nonlinear dependences.

Measuring the dynamic cross-correlation and contagion effect between Brent crude and heating oil (US-Diesel) prices pre and during the Covid-19 outbreak

Claudio M. C. Inacio Jr and Sergio A. David

Abstract: The objective of this work is to analyze the price dynamics and the level of association between the Brent crude oil prices and Heating oil (HO), i.e., US-Diesel. The data series are obtained from daily futures contracts prices of Chicago Mercantile Exchange (CME) group exchanges and Intercontinental Exchange (ICE). A continuous evaluation of the Detrended Cross-Correlation Analysis (DCCA) between Brent crude oil prices vis-a-vis HO is proposed by means of the rolling window approach allowing a dynamic analysis of their cross-correlations covering two periods, namely, from January/2018 to December/2019 (before the Covid-19 pandemic) and from January/2020 to December/2021 (during the Covid-19 scenario). The results indicate that there is a strong evidence of contagion in cross-correlation due to the initial impact of pandemics, but HO-Brent correlation fully recovered after approximately 200 days. However, lower time scales (n) are also sensitive to supply shortages in the short-term and can be most reliable for agents that might not take long positions. Measuring this dynamic cross-correlation can be a useful information for investor and agents in the oil and energy markets.

Forecasting Macroeconomic Tail Risk with Big Data Quantile Regressions

Jan Prüser and Florian Huber

Abstract: Modeling and predicting extreme movements in GDP is notoriously difficult and the selection of appropriate covariates is key in obtaining precise forecasts. In this paper, our focus is on using huge datasets in quantile regression models to forecast the conditional distribution of US GDP growth. To capture possible non-linearities between output and the regressors we include non-linear transformations of the data. This, in combination with the fact that, by definition, the number of observations in the tails is rather small, we propose a Bayesian approach to shrink our large-scale model towards a simpler model. Since Markov chain Monte Carlo

simulation becomes impractical in these dimensions, we rely on fast Variational Bayes approximations to the posterior distribution of the coefficients and the latent states. Our results suggest that huge information sets in combination with priors that introduce substantial shrinkage pays off for forecasting the tail risk of GDP.

Improve the forecasting accuracy using GJR-GARCH and MODWT (C6)

Tariq Alshammari, Mohd Tahir Ismail, Sadam Alwadi, Mohammad H. Saleh and Jamil Jaber

Abstract: This empirical research aims to model and improve the volatility pattern's forecasting accuracy by employing the Saudi Arabia stock market (Tadawul). This study uses daily closed price index data from October 2011 to December 2019 with the number of observations is 2048. In order to achieve significant results, this study employs many mathematical functions which are non-linear spectral model, Maximum overlapping Discrete Wavelet Transform (MODWT) based on autoregressive integrated moving average (ARIMA) model and Glosten-Jagannathan-Runkle-Generalized Autoregressive Conditional Heteroscedastic approach (GJR-GARCH) models. The present study's significant findings are that the entire previous event during the mentioned period of time will be explained while a new forecasting model will be suggested by combining the best MODWT using Coiflet function (C6) and the fitted GRJ-GARCH model. The performance of the forecasting will be evaluated based on some mathematical criteria such as Mean absolute percentage error (MAPE), Mean absolute scaled error (MASE) and Root means squared error (RMS).

Using Multi-Layer Perceptron (MLP) to Detect Outliers in Saudi Arabia Stock market

Khudhayr Rashedi, Mohd Tahir Ismail, Sadam Alwadi, Jamil Jaber and Nawaf Hamadneh

Abstract: This study aims to detect outlier of Saudi Arabia stock exchange (Tadawul) using the daily stock price indices data with 2026 observations from October 2011 to December 2019. This study employs multi-layer perceptron model for detecting outlier of stock returns classification. We have selected oil price (Loil) and repo rate (Repo) as input values according to multicollinearity test. The input variables in this study have been collected from Saudi Authority for Statistics and Saudi Central Bank. The output variable is obtained from Tadawul. The performance of MLP model is evaluated in terms of false positive rate (FP rate), false negative rate (FN rate), Fmeasure, Matthews correlation coefficient (MCC), and accuracy (ACC). The obtained results show that the performance of MLP is efficient based on different criteria tests.

The bootstrap for testing the equality of two multivariate stochastic processes with an application to financial markets

Ángel López-Oriona and José A. Vilar

Abstract: The problem of testing the equality of generating processes of two multivariate time series is addressed in this work. To this aim, we construct four tests based on a distance measure between stochastic processes. The metric is defined in terms of the quantile cross-spectral densities of both processes. A proper estimate of this dissimilarity is the cornerstone of the proposed tests. The first test employs the asymptotic distribution of the estimate, which we derive from some standard results on complex random variables. The low computational efficiency of this test makes it quite useless in practice. The three remaining techniques are based on the bootstrap. Specifically, a particular bootstrap method for spectral densities and extensions of the moving blocks bootstrap and the stationary bootstrap are used for their construction. The approaches are assessed in a broad range of scenarios under the null and the alternative hypothesis. The results from the analyses show that the procedure based on the stationary bootstrap exhibits the best overall performance in terms of both size and power. The proposed techniques are used to answer the question about whether or not the dotcom bubble crash of 2000s permanently impacted the global market behavior.

Bootstrap-based nonlinear shrinkage of sample eigenvalues

Eran Raviv

Abstract: Eigenvalues estimation is at the heart of many important statistical methods. Due to sample variability the sample eigenvalues are more dispersed than their population counterparts. This paper introduces a practical way to correct for the overdispersion using a nonlinear shrinkage function which is estimated by means of bootstrap. The algorithm proposed is intuitive, flexible, requires little in the way of hyperparameter tuning, and therefore computationally convenient. Experiments support the approach, both on a simulation study and two real-world financial data sets with recent data: US yield curve and industry portfolio returns. We estimate the importance of the first few factors, measured as usual by the proportion of explained variance. After applying the proposed bias-correction procedure we find that the first few factors are appreciably less influential than they may traditionally estimated to be, which carries important implications for downstream tasks, such as choosing the number of factors, or assessing the degree of commonality in the data.

Forewarning pest attacks in crops using beta regression

Bishal Gurung, Kn Singh, Achal Lama and Biwash Gurung

Abstract: Regression analysis is commonly employed to analyse data that are thought to be related to each other. It has to be one of the most used as well as one of the most misused statistical techniques. However, this technique may not be appropriate in statistical analysis of proportions as the proportions are restricted to the interval (0, 1). This restriction presents numerous difficulties, first one is it violates the assumption that the error term is normal and another one being the error term has constant variance. Many researchers have recognized this issue of analysing proportional data (Bartlett, 1936; Sokahl & Rohlf, 1995) and these researchers also worked on several analysis strategies to tackle the problem of proportions. To this end, for proportions that are obtained from discrete counts, logistic or binomial regressions are appropriate techniques (Quinn & Keough, 2002). For proportions that are not obtained from counts, there is still no concrete agreement on the most suitable techniques. A very well used recommendation is to apply a data transformation and proceed with ordinary linear regression (Crawley, 2012; Quinn & Keough, 2002; Sokahl & Rohlf, 1995). A major drawback of it is the difficulty in interpretability and the validity of the resulting inference. A number of researchers have identified different modelling strategies, which could be employed to model continuous random variables that assume values on (0,1) (Maier, 2014; Cribari-Neto & Zeileis, 2010; Hijazi & Jernigan, 2009; Ferrari & Cribari-Neto, 2004; Kieschnick and McCullough, 2003). These researchers have developed techniques to model continuous proportions that are easier to interpret and more flexible than transformation. To model proportions, some have employed beta regression and some have employed Dirichlet regression. In Beta regression the dependent variable follows a beta distribution and the mean of the distribution is related to a set of regressor variables through a linear predictor with unknown parameters and a link function. It also includes a dispersion parameter. For real data the normality assumption is violated often. The flexibility of the Beta distribution enables us to accommodate such skewed data. A huge amount of data on proportion are collected in agriculture too, so it becomes imperative to study the various possible applications and extensions of these methods for analysing such proportional data obtained from continuous observations. Data on percentage incidence of Helicoverpa armigera, Cucumo virus, Early blight, Ralstonia wilt, Sclerotium wilt, Tomato leaf curl, and Tomato Mosaic in different districts of West Bengal will be employed for analysis. Beta regression will be employed for the same.

The WEF's Global Competitiveness Reports: A case of reliable nonsense?

Harald Bergsteiner

Abstract: The World Economic Forum (WEF) ranked the US No. 1 in its 2018 Global Competitiveness Report and concluded that "The United States is the closest economy to the frontier, the ideal state" (WEF 2018:5). How did the WEF arrive at this conclusion, is the conclusion valid, and, does it matter? In this paper, I examine these and other questions. Every year the WEF ranks around 140 countries on a range of so-called competitiveness criteria. While the results reported by the WEF are relatively stable over time and so ostensibly meet reliability tests, the entirety of its country competitiveness work fails validity tests because only around 10 percent of the indicators actually measure what they purport to measure, namely country

performance/productivity/competitiveness. What the WEF actually measures is the extent to which countries comply with the WEF's untested input prescriptions for competitiveness. This mostly prescriptive rather than analytic approach enables the WEF to direct its research and publishing agenda to strongly-held ideological positions. In this paper, I therefore: explain why this matters; expose the ideological and unscientific nature of the WEF's country rankings; categorize the WEF's indicators according to whether they are prescriptive inputs or resulting outputs; provide guidelines on what scientifically valid country rankings should and should not do; and show that the WEF's prescriptions fail to deliver their asserted benefits.

Probabilistic Forecasting for Oil Producing Wells using Seq2seq Augmented Model

Hadeel Afifi, Mohamed Elmahdy, Motaz El Saban and Mervat Abu-Elkheir

Abstract: Time-series forecasting has been a challenging problem in the field of data mining. Deterministic forecasting has shown limitations in the field. Therefore, researchers are now more inclined towards probabilistic forecasting which has shown a clear advantage by providing more reliable models. In this paper, we utilize seq2seq machine learning models in order to estimate prediction intervals (PIs) for the large oil production dataset. To evaluate the proposed models, Prediction Interval Coverage Probability (PICP), Prediction Interval Normalized Average Width (PINAW) and Coverage Width-based Criterion (CWC) metrics are used. Our results show that the proposed model can reliably estimate PIs for production forecasting.

Geothermal time series analysis of the Deception volcano 2011-2022 (Antarctica)

Belen Rosado, Luis Miguel Peci, Amós De Gil, Paola Barba and Manel Berrocoso

Abstract: Deception Island is one of the most active volcanoes in Antarctica. It is loca-ted at the beginning of the expansion axis of the Central Bransfield Basin, between the Antarctic Peninsula and the South Shetland Islands, a geodyna-mically very complex region. The most recent eruptions date from 1842, 1912, 1917, 1967, 1969 and 1970. During the explosive eruptions occurred in 1967, 1969 and 1970, ash fall and lahars destroyed the Chilean and Bri-tish scientific bases, and significantly changed the island geomorphology. At the surface, the main indications of current volcanic activity are the existen-ce of fumarolic areas with gas emissions and termal anomalies of up to 100°C in the area known as Cerro Caliente. Since 1991, Deception Island is being mostly monitored by recording seismic activity and ground displacements. In 2011, a thermometric station was installed to study the temperature of the Cerro Caliente soil, in order to determine the state of volcanic activity on the island by correlating all these parameters. This work focuses on the analysis of the time series obtained from the thermometric station installed on Deception Island, from 2011 to 2022. A mathematical procedure is established that allows modeling the behavior and extracting the relevant information from the time series. This methodology consists of an initial treatment of the series in order to eliminate anomalous values that disturb the subsequent adjustments. A sequence of analytical fil-ters (Kalman, Wavelet), statistical filters (ARMA, ARIMA), and adjustment and forecasting techniques (CATS, STL) are applied to these filtered series. The results show that the periods of greatest surface deformation obtai-ned with GPS coincide with the increase in soil temperature recorded on Ce-rro Caliente. A significant event is observed in the 2012 austral winter, ini-tially detected in the seawater temperature and in the soil temperature on Ce-rro Caliente. These anomalies were precursors to the other analyzed obser-vables – the seismicity and ground deformation data – registered only in the following austral summer campaign, suggesting rapid and near-simultaneous increase in geothermal activity and the beginning of an inflation process on the island.

An analysis of liquidity through the self-similarity surfaces

Sergio Bianchi, Augusto Pianese and Massimiliano Frezza

Abstract: Scale invariance, also referred to as self-similarity, is the property characterizing those phenomena which preserve the same (physical or probabilistic) structure regardless the degree of magnification on a (state) space or time dimension. This property characterizes many stochastic processes: the (fractional) Brownian motion or the Bessel process, the α-stable L´evy motion (sLm), the generalized Hermite processes, which include the Rosenblatt process or other non Gaussian stochastic processes...

Analysis of underwater temperature series of Deception and Livingston Islands 2012-2022 (Antarctica)

Olga Luengo, Belén Rosado, Amós De Gil, Paola Barba and Manuel Berrocoso

Abstract: Deception and Livingston Islands belong to the Bransfield Strait, this region is a geodynamically complex area caused by the interaction of several tectonic plates. Deception Island is one of the active volcanoes in this subduction zone, so Livington Island is taken as a reference. The underwater temperature records in Antarctica are rare and most of them correspond to the antarctic summer due to the weather conditions during the rest of the year and the formation of the ice sheet. After several attempts long continous records were obtained in both islands from 2012 until today. The series of both islands were studied using conductivity-temperature-depth sensors for about 10 complete years. Temperature records range from -2 to 2ºC, highest temperaturas occurs during the austral summer or during periods of volcanic activity in the case of Deception Island, on the other hand, lowest temperatures recorded during the austral winter. During the austral winter, sometimes because of the low temperaturas, ice sheet are formed. The aim of this work is to study the temperature periodicities and to compare them between the two islands in order to identify the differences that may be caused by the influence of volcanic activity. This is achieved by using statistical and analytical techniques for time series analysis.

Analysis of tidal series of Deception and Livingston Islands (Antarctica)

Olga Luengo, Belén Rosado, Amós De Gil, Paola Barba and Manuel Berrocoso

Abstract: Deception and Livingston Islands are part of the South Shetland Island chain, this archipelago form the northen boundary of the Bransfield Strait and the Antarctic Peninsula forms the sourthern bourandy. Tide records are highly valued fot tide forecasting in the región for scientific work. However, it is difficult to install and maintain equipment on the Antarctic continent and its vicinity due to the adverse conditions of the austral winter. Despite this, two tide stations were established in 2012, one at Deception Island (DECMAR) and other at Livingston Island (LIVMAR). Eventually continuous records have been recorded since 2015 in Livingston and 2016 in Deception. These time series, were studied using conductivity-temperature-depth sensors with a sampling inverval of 20 min. The tidal constituents are solved by a yearly harmonic analysis with results about 98% agreement with the predictions stimated. Tides in the Bransfield Strait are described as mixed, mainly semi-diurnal, so there are daily two high and two low waters. A detailed study of the influence of periodicities (semi-diurnal, diurnal, fortnightly, monthly, semi-anual and anual) on the calculation of tidal costituents is carried out in this work.

Influence of the Eruption of La Palma on the Evolution of Tropospheric Activity in the Canary Archipelago

Paola Barba, Belén Rosado, Javier Ramírez-Zelaya, Adrián Villanueva and Manuel Berrocoso

Abstract: The signals emitted by the GNSS satellites propagate through a dispersive medium such as the atmosphere until they are received at the observation station, producing a delay in the reception of said signal. This delay, due to the medium in which the signal propagates, can be considered as the composition of the ionospheric and tropospheric effects. By estimating these delays, it is possible to use gnss for high-precision positioning but also as a valuable tool to monitor the troposphere. Zenith total delay (ZTD) or precipitable water vapor (PWV) parameters are considered. The tropospheric correction intrinsically depends on the meteorological conditions existing at the station at the moment of the observation, being necessary to apply mathematical models that eliminate or reduce this effect in the observations made. The contribution of volcanic material that is generated during volcanic eruptions to the atmosphere causes the atmospheric models obtained to present large anomalies and deviations with respect to the existing tropospheric situation in the absence of these processes. The ZTD and PWV parameters reflect the evolution of the intensity of the eruptive process both spatially and temporally. In this work, the spatial and temporal evolution of the eruptive process that occurred on the island of La Palma (Canary Islands) from September 19 to December 13, 2021 is studied. The time series of the PWV parameter obtained from the GNSS observations obtained during the years 2020, 2021

and 2022 in continuous monitoring stations and distributed throughout the Canary Archipelago are established. These series are analyzed using statistical (ARMA, ARIMA) and analytical (Kalman, wavelets) techniques programmed in R.

Economic and Legal Nature of Blockchain

Farouq Alazzam and Iryna Diachuk

Abstract: Abstract. The dynamic development and introduction of advanced technologies have led to the natural transformation of social relations. Thus, the rapid development of blockchain technology as a conceptually new type of resource accounting and ensuring the highest degree of transparency, security and trust, as well as the active integration of this technology into various fields of activity, served as the starting point for the formation of a new type of socio-economic relations, the regulation of which should be carried out in the legal plane. All of the above testifies in favor of the need to create a legislative framework in order to provide legal support for a new type of socio-economic relations. In this regard, there was a need to consider the mechanism of operation and the principles on which the blockchain technology is based, for deep and correct understanding of the essence and practical usefulness of this technology. Despite the active introduction of blockchain technology and blockchain tokens in various areas of public life, their capabilities and benefits have not been sufficiently studied yet. There are discussions among theorists and practitioners about the prospects for using blockchain technology in the financial, banking and social sectors, as well as in the public one. The article is devoted to a complex analysis of blockchain technology and new social relations arising from the use of products and services based on this technology. The article describes the economic and legal nature of the blockchain token; discloses the economic, technical and legal components of the balance sheet of digital assets, which allows keeping records of any values using blockchain technology and calculating the main economic performance indicators. The meaning of the concepts of "anonymity", "identification" and "verification" is determined on the basis of a specific approach: by distinguishing between the concepts of anonymity, definition and classification, it is argued by applying the correct norms that oblige various platforms and services based on a sample of user data processing. Blockchain tokens that have the above-described feature - provide the right to access a certain set of rights to a value, can be compared with such a category as a digital asset - an information resource derived from the right to a value and circulating in a distributed ledger in the form of a unique identifier. Thus, a digital asset by means of an identifier provides its owner with the right to access an information resource containing a set of rights (rights of ownership, use and disposal) for value (property, property and non-property rights, securities, services, works, digital assets and other values). So, blockchain tokens in terms of their economic and legal nature can be classified as a speculative asset or a digital asset. The use of a digital asset allows accounting for a set of rights to a value (asset) through software solutions based on the use of distributed ledger technology. Special attention in the article is paid to the issue related to the balance sheet accounting of blockchain tokens. The balance sheet of blockchain tokens is defined as a method of registering and recording transactions implemented by the blockchain network node software, which allows to determine the difference between receipts and expenditures of accounting units at a certain point in time at their accounting address. It is worth mentioning that one of two accounting models for blockchain tokens is currently used in distributed ledger networks: UTXO or Account Based Model. Thus, the use of digital assets allows the subjects of legal relations to keep records of any values using blockchain technology, as well as to calculate the main economic indicators of activity, in particular: the volume of output of goods, the cost of production, profit and profitability of activities. References 1. A Guide To Digital Token Offerings. URL: https://www.mas.gov.sg/-/media/MAS/ Regulations-and-Financial-Stability/Regulations-Guidance-and-Licensing/Guideto-Digital-Tokens-Offering-last-updated-on-5 April-2019.pdf. 2. A Securities Law Framework for Blockchain Tokens. URL: https:// www.coinbase.com/legal/securities-law-framework.pdf. 3. A Statement on Launching «Aber» Project, the Common Digital Currency between Saudi Arabian Monetary Authority (SAMA) and United Arab Emirates Central Bank (UAECB). URL: http://www.sama.gov.sa/en-US/News/Pages/news29012019. aspx. 4. Act To Amend The Estates, Powers And Trusts Law, In Relation To The Administration Of Digital Assets. URL: https://legislation.nysenate.gov/pdf/ bills/2015/A9910A. 5. ADGM successfully concludes 1st industry E-KYC utility project with UAE financial institutions. ADGM FSRA. 04.12.2018. URL: https://www.adgm.com/ media-center/announcement-listing-page/media-releases/adgm-successfullyconcludes-1st-industry-e-kyc-utility-project-with-uae-financial-institutions/. 5. Bitcoin Is Not a Virtual Currency. GERMAN BUNDESBANK. Feb. 20, 2018. URL: https://www.bundesbank.de/

Redaktion/EN/Topics/2018/2018_02_19_diskussion_ bitcoin.html?https=1. 6. Blockchain and GDPR. URL: https://www.eublockchainforum.eu/reports. 7. Blockchain for digital government, EUR 29677 EN. Publications Office of the European Union. Luxembourg, 2019, ISBN 978 92 76 00581 0, doi: 10.2760/942739, JRC115049. URL: https://joinup.ec.europa.eu/sites/default/files/document/2019–04/ JRC115049%20blockchain%20for%20digital%20government.pdf. 8. Blockchain for government and public services. URL: https://www.eublock chainforum.eu/reports. 9. Blockchain in Public Sector Transforming government services through exponential technologies. Deloitte, FICCI. January 2018. URL: https://www2.deloitte.com/ content/dam/Deloitte/in/Documents/public-sector/in-ps-blockchain-noexp.pdf. 10. Blockchain innovation in Europe. URL: https://www.eublockchainforum.eu/ reports. 11. Blockchain Investment Trends (1H 2019). TeqAtlas. Aug 1, 2019. URL: https:// teqatlas.com/analytics-and-research/fs776 blockchain-investment-trends-1h-2019. 12. Blockchain technology and decentralized governance: Is the state still necessary? Journal of Governance and Regulation.№6(1).P.45–62. URL: http://doi.org/10.22495/ jgr_v6_i1_p5. 36. Blockchain World Wire. URL: https://www.ibm.com/blockchain/solutions/worldwire. 13. BOJ/ECB Joint Research Project on Distributed Ledger Technology. ECB & BOJ, Sept. 6, 2017. URL: https://www.ecb.europa.eu/paym/intro/news/shared/20170906_ stella_report_leaflet.pdf.

Impact of capitalization on labor productivity in European countries

Antonio Mihi-Ramírez, Elías Melchor-Ferrer and Ammar Ziad-Alwrekiat

Abstract: Advances in European integration in recent decades have resulted in a high and growing interdependence between European nations in terms of both product markets and production factors and are highly outsourced. These facts are essen-tial when explaining the development of labor productivity in these countries giv-en the high mobility in the capital factor and in the labor force, which directly af-fects it. This study aims to analyze the extent to which physical and human capital affect the productivity of the labor factor in a group of European countries be-longing to the EU (Austria, Belgium, Denmark, Finland, France, Germany, Ire-land, Italy, the Netherlands, Portugal, Spain, Sweden, and the United Kingdom) during the period 2000-2015. To do this, as proxy variables for physical and hu-man capital we will take the capital stock index in services and Employment rates by place of birth and educational attainment (25-64), respectively, in both cases, with data obtained from Eurostat. Given the high interrelationship between these countries and the mobility of pro-ductive factors, it is foreseeable that spatial overflows will be observed, so that what happens in a country in terms of increasing or decreasing returns in the use of said factors may affect the endowment of these for their neighbors and, conse-quently, in the labor productivity observed. However, this analysis implies the se-lection of the most appropriate spatial weights matrix. In our case, we have con-sidered several specifications based on geographical distance: k-nearest neighbors (k = 1 to 10), distance between the centroids of the various regions from mini-mum to maximum and with cut-off kilometers for intermediate distances, and in-verse distance matrices with a distance decay factor whose off-diagonal elements are defined by $wij = 1/dij\alpha$, with α = 1.00, 1.50, … and 3.00. The starting point is a Spatial Durbin Model (SDM) with spatial and period fixed effects. The inter-est in spatial econometrics literature on the estimation of econometric relation-ships based on spatial panels can be explained by the extended modeling possibil-ities for researchers. In our case, we have adopted the “two nearest neighbors” matrix, not only because the estimation procedures indicate it, but also because we are dealing with only thirteen territorial units that, moreover, are countries. An exploratory spatial analysis of the data through the Moran Index allows us to verify the existence of a weak positive autocorrelation for the capital stock in ser-vices, while for labor productivity it presents a negative autocorrelation with a significance of 5%. This last circumstance is also observable in some cases for employment rates by level of educational attainment, specifically, for those born abroad and very particularly for medium and low educational levels, which would indicate two things: on the one hand, the greater mobility of the foreign work-force compared to that of national origin, and on the other, that the population with a higher level of education is more on the margin of this type of dynamic. Likewise, if we obtain the LISA cluster maps or quartile maps, we can observe a spatial association for the aforementioned variables. Once we have confirmed the existence of spatial dynamics in both the explained variable (labor productivity) and the explanatory ones (employment rates by na-tionality and level of educational attainment, and capital stock in services), we can select that spatial econometric model that best suits this phenomenon under study. For this, the starting point is going to be the general nesting spatial model (GNS), from which other different models are derived depending on the inclusion of cer-tain spatially lagged variables: spatial Durbin model –SDM–, spatial Durbin error model –SDEM–,

spatial autoregressive combined model –SAC–, spatial auto-regressive model –SAR–, spatial error model –SEM–, spatial lag of X model –SLX–, or ordinary least squares –OLS–). The classification of the main spatial econometrics models is based on the three types of spatial interaction: i) an en-dogenous interaction when the economic decision of an agent or geographical zone will depend on the decision of its neighbors; ii) an exogenous interaction when an agent's economic decision will depend on the observable characteristics of its neighbors, and iii) a spatial correlation of the effects due to the same unob-served characteristics. Along with the introduction of spatial heterogeneity and dependence in the model, to determine the best estimation strategy we have also decided to use panel data models. This choice opens the possibility of specifying what type of effects are, a priori, observable (random or fixed and, within the latter, spatial and/or time-period). After applying the Hausmann test, random effects are excluded in favor of fixed ones, in which both spatial and temporal effects are significant. Regard-ing the model selection procedure, as usually in the literature, we have chosen to start from the SDM model since its estimators are unbiased, which makes it more robust in the case of poor specifications. In addition, two others are derived from this model, the SAR model and the SEM model based on the likelihood ratio test results. In both cases, the SDM model cannot be simplified in either SAR or SEM, so the latter two would be discarded. The alternative procedure to estimate the models is the Bayesian method, which produces posterior model probabilities and examined the spatial regression mod-els. The probabilities for these models sum to unity since the different weight ma-trices are compared for each model (the higher the value, the better the perfor-mance). However, the results are similar to those commented on previously. The estimation of the SDM models has well-adjusted goodness of fit, regardless of the level of educational achievement considered and the nationality of birth of the workers, since the coefficient of determination is in all cases around 98%, alt-hough if we discount the effect of the fixed effects the corrected coefficient of de-termination ranges between 37 and 46%, which indicates the existence of com-mon spatial and temporal patterns with an apparent greater explanatory power than the variables themselves. If we add that in all cases the spatially lagged labor productivity is not significant, it leads us to rethink the selection of the model. To the extent that the explanatory variables (with or without spatial lag) are generally significant, we have chosen to use an SLX model without the presence of fixed effects to determine the isolated impact of the capital stock in services and the employment rates by nationality of birth and level of educational attainment on la-bor productivity. Regarding the capital stock in services, from the analysis of the estimates made, it can be concluded that both this and its spatial lag are significant in all cases, but with different signs (positive and negative, respectively), which implies that the regions with high levels of capital stock present higher levels of labor productivi-ty, while those same high levels for neighboring regions would have a negative impact on the region's productivity, which could be explained by the existence of increasing returns of scale in neighboring regions that would attract investment, activity, employment, etc. On the other hand, a more detailed analysis of the value of the estimated coefficients enables us to verify that in most cases the indirect impact (of neighboring regions) of the capital stock in services on labor produc-tivity offsets the direct and positive impact of the region itself. The only excep-tions would correspond to employment rates for high educational attainment lev-els, regardless of their nationality of birth. As regards the employment rates by level of educational attainment, the estimates obtained allow us to conclude that for the high educational level these rates do not have explanatory power on labor productivity in any case. On the other hand, for medium and low educational levels, the relationship with labor productivity would be significant, but with a negative sign, that is, it would tend to offset the positive effect of the capital stock on labor productivity. The level of educational attainment in neighboring regions does not seem to be significant in the level of labor productivity, only and with low significance for the whole of the population born abroad, but because the same occurs for the level low education. That is, when neighboring regions have a high employment rate for the foreign-born population with a low educational level, this has a negative impact on the labor productivity of the region itself, which would feed back the negative direct effect of such a level of education in the region. From a methodological point of view, one of the main problems faced by a re-searcher trying to approach econometric modeling from a spatial point of view is mastering the possible techniques and options. Although there are multiple op-tions in different statistical packages, in our case we have chosen to use the MATLAB software given the abundance of freely available functions. In addi-tion, we have integrated into a single file all the necessary steps and options to consider in the spatial econometric modeling process: i) Selecting the spatial weight matrix. Through an iterative process, a Spatial Dur-bin Model (SDM) with spatial and time-period fixed effects is estimated for all spatial weighting matrices considered. The matrix gives a much better fit than others (its residual variance value is smaller than other matrices) will be selected. ii) OLS without fixed effects, spatial effects, time-period effects, spatial and time-period effects, and random effects; in all cases with LM tests for spatial lag and spatial error model fixed

effects, within estimator. Moreover, the additional test considers the (null) hypothesis that the spatial (and time-period) fixed effects are jointly insignificant must be rejected. iii) To estimate the direct/Indirect effect of the explanatory variables, including bi-as correction of coefficient estimates by Lee and Yu (2020), as well as a selection framework to determine which spatial panel data model best describes the data (SDM, SAR, or SEM) and the significance of effects (fixed spatial and/or period, or random). iv) To estimate OLS, SAR, SEM, SLX and SDEM. Due to the Monte Carlo character of the different estimations, every time this option is run again, the re-sults (and the spatial parameters) will be slightly different. v) Comparing static panel data models with dynamic spatial panel data models with fixed or random effects. Free and open-source software development practices have given rise to a new vision of the construction, deployment, and evolution of complex software sys-tems. There are great benefits to the release and licensing of spatial econometric code under the GNU Public License, with the source code freely available for study and code improvement. For all these reasons, our source code compilation is being prepared for downloaded for free.

LASSO regression: an application for Brazilian hydrological time series

Johann Freitas

Abstract: Hydreletric energy is the most important source in Brazilian power generation system. Water resources intertemporal management plays a crucial role on minimizing costs for the operator, the whole Economic chain, the final costumer, and on long term system sustainability. Then, urges the need for models capable of predicting future riverflows accurately. This study aims to evaluate Machine Learning, in particular the LASSO regression, applicability for Brazilian Monthly Riverflow series, in terms of estimates feasibility and forecast accuracy against traditional identification methods. Since this type of data can be represented using a PAR approach, due to it's seasonal behaviour, LASSO emerges as an alternative for defining a vector of autoregressive orders. The result shows that LASSO can automatically return a physical consistent representation for data, without losses in terms of predictive performance.

Forecasting Net Asset Value (NAV) for Exchange Traded Fund(ETF) using Time Series ARIMA Model

Zobia Mushtaq and Muhammad Rafi

Abstract: Exchange Traded Fund (ETF) is an innovative financial investment instrument. It combines features and benefits of stocks, mutual funds and bonds. ETF market has grown rapidly in a short span of time all over the world. In Pakistan, it is recently introduced and it popularity is increasing day by day. The paper reports the performance and forecasting of Meezan Pakistan Exchange Traded Fund (MZNPETF)’s Net Asset Value (NAV). This ETF grow more than 30% in last 1.5 years and the payout for this funds was 1.25 per unit last year. The study uses daily NAV of the fund from 7 October 2020 Till 10 March 2022. An autoregressive integrated moving average (ARIMA) which is a statistical analysis model that uses time series data to perform both performance analysis and forecasting is used in the study. ARIMA model is developed using first 80% of the data and tested on last 20% as a test-set. The adequate model was selected according to performance of four criteria of Akaike criterion, Schwarz Bayesian criterion, log likelihood, and standard error. The selected model is compatible with ARIMA (1,1,1).and as evaluation metric we achieved Root Mean Square Error (RMSE) value of 0.120, thus creating a very efficient forecasting model.

Economic uncertainty impact in forecasting Uruguayan macro variables

Bibiana Lanzilotta, Gabriela Mordecki and Viviana Umpiérrez

Abstract: Uruguay as a small and open economy is highly exposed to international, regional, and domestic uncertainty. This paper, using two alternative indicators explores the impact of the uncertainty shocks on nominal and real macroeconomic variables, and their capability to anticipate macroeconomic fluctuations between 1997 to2021. The main results suggest that economic uncertainty has a significant impact on the real economy (in the industrial production), while there is less evidence of impact on financial variables in Uruguay. This result remains unchanged despite the large Covid-19 shock. Moreover, both indicators anticipate cyclical fluctuations, and their predictive performance does not show significant differences, with the best predictor

varying at different horizons. Nevertheless, we find evidence that correlations between the cycle and these uncertainty indicators vary along the analyzed period, with higher, showing higher values (in absolute terms) during an economic crisis.

A decoder-based transformer model for multi-step fresh produce price forecasts

Fatemeh Jafari, Kumaraswamy Ponnambalam and Fakhri Karray

Abstract: This study investigates the efficiency of a multivariate decoder-based transformer model for multi-step fresh produce (FP) price forecast up to seven days ahead. The proposed model is compared with two baseline approaches, i.e., a simple LSTM deep network and an SVR conventional ML model, in terms of accuracy and efficiency. The daily yield and price datasets of strawberries related to Santa Maria, California, between the years 2011 to 2019 are used for the experimentation. Results show that the investigated deep network models improve the forecast accuracy by up to 37%, compared to the SVR approach, especially for the larger forecast horizons. Moreover, the transformer performs more efficiently than LSTM, with about twelve times faster convergence. Besides, using yield as an additional predictor in the multivariate FP price forecast enhances the accuracy by up to 14% compared to the univariate version of the model, which uses the price as the only input.

K-means Clustering Assisted Spectrum Utilization Prediction with Deep Learning Models

Bethelhem S. Shawel, Frehiwot Bantigegn, Tsegamlak Terefe Debella, Sofie Pollin and Dereje H. Woldegebreal

Abstract: The radio spectrum is a finite and scarce resource needed to transport data generated by existing and emerging wireless mobile networks and services. As the demand for wireless services is increasing, operators look for ways to efficiently utilize their assigned spectrum. While operators do regularly perform spectrum occupancy measurement using an external spectrum analyzer or installing a dedicated sensing network to understand and plan the spectrum utilization level, both in time and spatial dimensions, such a measurement-based approach is expensive, given the dynamic and wide range covering nature of spectrum utilization. This paper proposes an indirect approach to assess and predict the average spectrum utilization level using data traffic measured from base stations of an operator network. K-means clustering and deep learning algorithms; namely, Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM), are used to model and analyze the current and future spectrum utilization in the 900 MHz frequency range. Data collected from 639 base stations are used to build the spectrum utilization model. The results show that the CNN model trained on clustered data outperforms the model developed on non-clustered data (with an RMSE of 0.58), mainly for base station level pre-diction. In terms of utilization level, the results also show that the operator is not optimally utilizing the 900 MHz range.

A Multivariate Approach for Spatiotemporal Mobile Data Traffic Prediction

Bethelhem S. Shawel, Endale Mare, Tsegamlak Terefe Debella, Sofie Pollin and Dereje Hailemariam Woldegebreal

Abstract: A widespread deployment of spectrally efficient mobile networks, advancements in mobile devices, and proliferations of attractive applications has led to an exponential increase in mobile data traffic. Mobile Network Operators (MNOs) benefit from the associated revenue generation while putting efforts to meet customers' expectations of delivered services. Having a clear knowledge of the traffic demand is critical for network dimensioning, optimization, resource allocation, market planning, and the likes. As the traffic demand, among others, is a function of customers' behavior, settlement pattern, land use, and time of a day, capturing the traffic characteristics both in temporal and spatial dimensions is needed. Moreover, other parameters such as the number of users and data throughput inherently contain traffic-related information, necessitating a multivariate approach to understand the traffic demand. Realizing the multidimensional and multivariate nature of data mobile data traffic, in this paper we propose a multivariate and hybrid Convolutional Neural Network and Long Short Term Memory network (CNN-LSTM) data traffic prediction model. The model is built based on mobile traffic data collected from the Long Term Evolution (LTE) network of an operator network. The results confirm that the proposed model outperforms its univariate counterpart in Root Mean Square Error

(RMSE) and Mean Absolute Percentage Error (MAPE) by 58% and 50%, respectively. Moreover, the model is further compared with CNN only univariate and multivariate models, and it outperforms the two models. The comparisons substantiate the achievable improvements because of the hybrid and multi-variate nature of the prediction algorithm.

Comparative analysis of residential load forecasting with different levels of aggregationComparative analysis of residential load forecasting with different levels of aggregation

Ana Apolo Penaloza, Roberto Leborgne and Alexandre Balbinot

Abstract: Microgrids need a robust residential load forecasting. As a consequence, this highlights the problem of predicting electricity consumption in small amounts of households. The individual demand curve is volatile, and more difficult to fore-cast than the aggregated demand curve. For this reason, Mean Absolute Per-centage Error (MAPE) varies in a large range (of 1% to 45%), depending on the number of consumers analyzed. Different levels of aggregation of household consumers that can be used in microgrids are analyzed, The load forecasting of the single consumer and aggregated consumers are compared. The forecasting methodology used is the most consolidated of Recurrent Neural Networks, i.e., LSTM. The data set used contains 920 residential consumers belonging to the Commission for Energy Regulation (CER), a control group that is in the Irish Social Science Data Archive (ISSDA) repository. The result shows that the forecasting of groups of more than 20aggregated consumers has lower MAPE that individual forecasting. On the other hand, individual forecasting is better for groups with less than 10 consumers

Ensemble Forecasting by Energy associated modified by Renyi's entropy and statistical roughness in the learning process

Cristian Rodriguez Rivero, Julián Pucheta and Martin Herrera

Abstract: This paper proposes an ensemble method based on Energy associated modified method by Renyi's entropy and statistical roughness to tune recurrent neural networks (E-EAS-LSTM) parameters. The computational models are carried out for univariate chaotic time series with multi-step prediction horizons. The combination of energy associated with series and the relative entropy combines heuristically modified the learning process. The study analyses and compares the relative advantages and limitations of several algorithms, including statistical methods to forecast both benchmark series and rainfall series from 1 to 6 months ahead. Simulation results illustrate that regardless of the effectiveness of the E-EAS-LSTM approach through different series classified by the proposed approach in both, the learning process and the validation test using the SMAPE and NRMSE, and MASE metrics, and by adding fractional Gaussian noise, with a good performance in terms of accuracy.

Feature Importance Study on Forecasting Hourly Retail Time Series

Bruno Mendes, Ana Maria Tomé and José Moreira

Abstract: This work presents a predictive linear approach to hourly retail time series. The preprocessing steps must cope with time schedule normalization. After model fitting, a feature ranking strategy is applied to study the relevance of the lags for the prediction. An experimental study was conducted with different stores and different periods for training and testing. This study resulted in a low-complexity and robust regression pipeline which can be applied to this type of time series.

Evaluating a Recurrent Neural Networks Model For Predicting Readmission to Cardiovascular ICUs Based on Clinical Time Series Data

Sobhan Moazemi, Sebastian Kalkhoff, Steven Kessler, Zeynep Boztoprak, Vincent Hettlich, Artur Liebrecht, Roman Bibo, Bastian Dewitz, Artur Lichtenberg, Hug Aubin and Falko Schmid

Abstract: Unexpected readmissions to intensive care units (ICUs) en- danger patients' lives due to premature patient transfers or prolonged stays at the care units and expose unnecessary costs to healthcare sys- tems. This can be mitigated by stratification of the readmission risk at discharge times using state-of-the-art machine learning (ML) based so- lutions. In this study, we evaluate a Recurrent Neural Network (RNN) model based on Long Short-Term Memory (LSTM) which was trained, vali- dated and tested using the Medical Information Mart for Intensive Care (MIMIC-III) dataset and further evaluated on an independent cohort of heart surgery related ICU stays in the time period from January 2017 to December 2020 at the Department of Cardiovascular Surgery at the University Hospital Düsseldorf (UKD). To train and fit the model, time series data of laboratory and vital values from a subset of the MIMIC-III dataset with 11513 cardiac surgery related ICU stays has been applied. The main objective of this study was to evaluate the performance of two alternative models for the prediction of readmission to the ICUs. The first model is trained and tested using all the available time series data from each patient's corresponding ICU stay, while the second model is trained and tested using the data from the last 48 hours of the ICU stay before being admitted to a normal station. The readmitted patient cohort consisted of patients which were either readmitted to the ICU within 48 hours after their discharge or transfer to a normal station as well as subjects who had died within 30 days after their discharge or transfer from the ICU. Patients who were not readmitted to the corresponding ICU and did not die within 30 days after discharge or transfer were used as control subject cohort and were labeled as non-readmitted. For our two evaluation sets 966 and 10547 patients from the MIMIC database are labeled as readmitted and non-readmitted respectively. Likewise, 100 and 402 patients from the UKD dataset are labeled ac- cordingly. The case feature vectors include 10 clinical and heart-related features (such as body temperature, heart rate, Sodium, and Bilirubin) which were identified as meaningful by cardiac surgeons and are sampled in 60 minute intervals throughout the patients' stays at the corresponding ICU as time series data. In addition the patients' age, weight, and length of stay (LoS) are extracted from the patients file. The recurrent models first analyze the time series data using a LSTM method, then concatenate the 3 patient-specific parameters (age, weight, and LoS) to the feature vectors. The results suggest the relevance of the time series based approach for readmission prediction with area under the curve of precision-recall (AUC-PR) of up to 0.63. Furthermore, the findings revealed the supe- riority of the model with the fixed time frame of 48 hours on the time series data to the alternative model with no fixed time frame as applied to the independent subject cohort from UKD. This would emphasise the positive impact of consistency in data driven approaches towards health forecasting. Moreover, it signifies the rele- vance of the patients' clinical factors measured in the most recent times- tamps prior to transfer in predicting readmission risks. To conclude, we have analyzed the predictive power of an ML based ap- proach for readmission stratification in cardiovascular patients staying at ICUs. The results suggest that the RNN model based on time series forecasting holds promise to be integrated in clinical routines as clinical decision support tools in the future. To achieve this goal, follow-up stud- ies with bigger cohorts of multicentric data as well as proper user studies should be conducted to further assess the generalizability and usability of the methods respectively.