DEUTSCHE BUNDESBANK
EUROSYSTEM

An overall seasonality test based on recursive feature elimination in conditional random forests

Karsten Webel, Daniel Ollech / Deutsche Bundesbank, DG Statistics

5th International Conference on Time Series and Forecasting, Granada, 20 September 2018

Agenda

- Motivation
- Random forests and conditional inference trees
- Candidate and overall seasonality tests
- Simulation algorithm and design
- Results
- Summary
- References

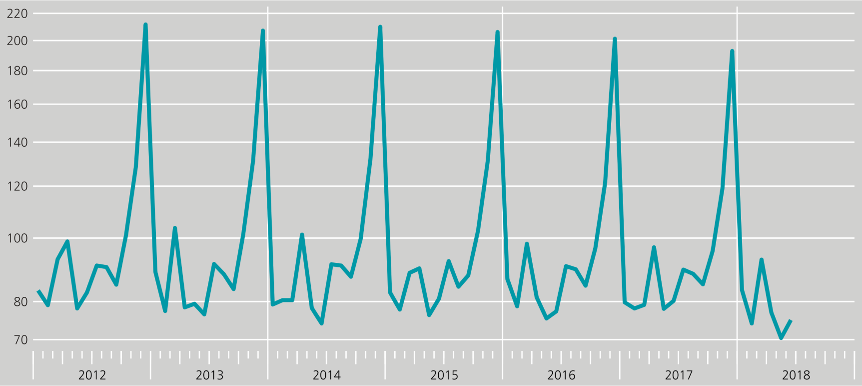
What is next?

- Motivation
- Random forests and conditional inference trees
- Candidate and overall seasonality tests
- Simulation algorithm and design
- Results
- Summary
- References

What is seasonality? (I/II)

Retail trade turnover: games and toys

Value, 2015 = 100, log scale



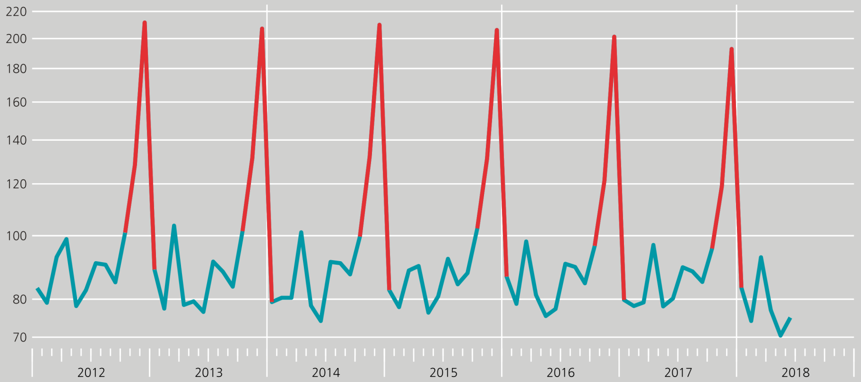
Deutsche Bundesbank

S3PR0323.Chart

What is seasonality? (I/II)

Retail trade turnover: games and toys

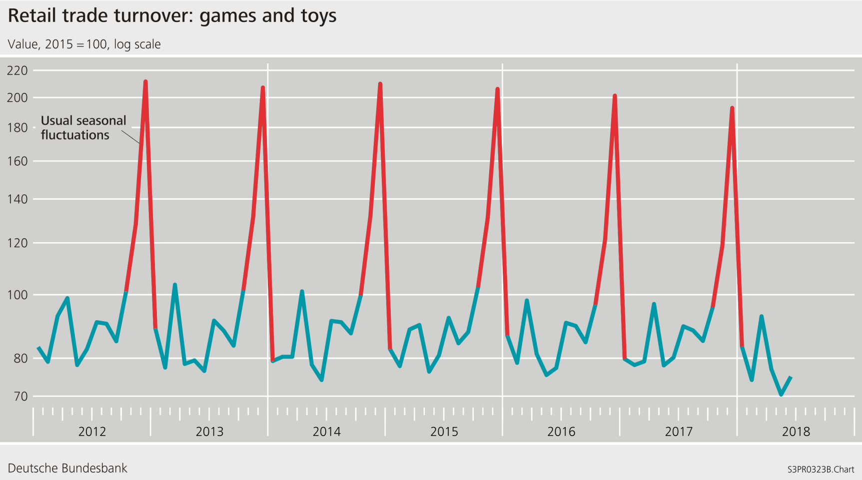
Value, 2015 = 100, log scale



Deutsche Bundesbank

S3PR0323A.Chart

What is seasonality? (I/II)



What is seasonality? (II/II)

Deutsche Bundesbank, Statistical Supplement 4
"Seasonally adjusted business statistics"

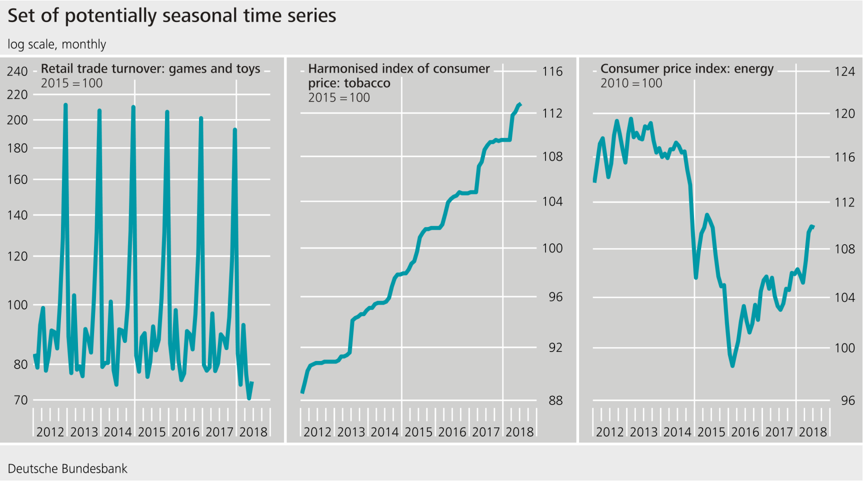
'**Usual seasonal fluctuations**' means those movements which

- ☞ **recur with similar intensity**
- ☞ **in the same season each year**

and which, on the basis of past movements of the time series in question,

- ☞ **can**, under normal circumstances, **be expected to recur**.

Are these economic time series seasonal?



Answers from JDemetra+

Are these economic time series seasonal?			
Test	Retail trade turnover: games and toys	HICP: tobacco	CPI: energy
QS	Yes	No	Perhaps
FT	Yes	No	Perhaps
KW	Yes	No	Yes
SP	Yes	No	No
PD	Yes	No	Yes
SD	Yes	No	Yes

Research questions

How can we combine results of a set of seasonality tests?

- ☞ Classification problem
- 💡 Random forest

How can we separate relevant from irrelevant seasonality tests?

- ☞ Informational content
- 💡 Variable importance measures

How can we construct an informative overall seasonality tests?

- ☞ Sequential “winners stay – loser walks” competition
- 💡 Recursive feature elimination

What is next?

- Motivation
- Random forests and conditional inference trees
- Candidate and overall seasonality tests
- Simulation algorithm and design
- Results
- Summary
- References

Random forest (I/III)
Notations

Training data

$$\mathcal{L} = (\mathbf{X}\mathbf{Y})$$

Predictors

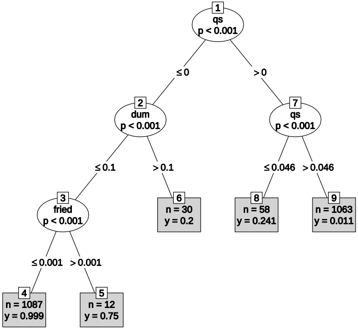
$$\mathbf{X} = (\mathbf{X}_1 \ \dots \ \mathbf{X}_p) \quad \text{with} \quad \mathbf{X}_j = (x_{1j}, \dots, x_{Nj})^\top$$

Categorical response

$$\mathbf{Y} = (y_1, \dots, y_N)^\top \quad \text{with} \quad y_i \in \{1, \dots, K\}$$

Random forest (II/III)
Basic idea

Classification tree

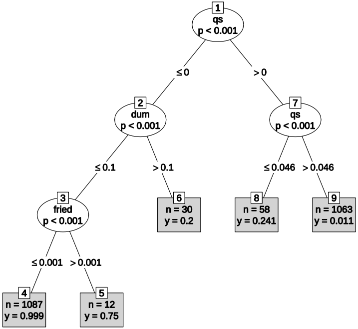


Single vote

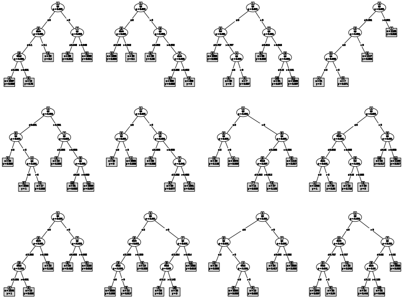
Random forest (II/III)

Basic idea

Classification tree



Independent classification trees



Single vote

Majority vote

Random forest (III/III)
Potential issues

Classical approach

- Variable selection, node splitting $\rightsquigarrow$ No separation
- Source $\rightsquigarrow$ Sampling scheme at each node

Potential biases

- Variable selection $\rightsquigarrow$ $\mathbf{X}_j$'s with larger measurement scales, higher number of categories, missing values
- Variable importance measures $\rightsquigarrow$ Correlated $\mathbf{X}_j$'s (in addition)

Potential consequences

- Truly influential $\mathbf{X}_j$ $\rightsquigarrow$ Underestimated importance
- Seemingly influential $\mathbf{X}_j$ $\rightsquigarrow$ Overestimated importance

Conditional inference trees (I/III)
Basic idea

Bootstrap samples (with replacement, $b \in \{1, \dots, B\}$)

Sampled ("bag") data	$\rightarrow$	$\mathcal{L}_b$	$\rightarrow$	Unpruned tree $\mathcal{T}_b$
"Out-of-bag" (OOB) data	$\rightarrow$	$\mathcal{O}_b = \mathcal{L} \backslash \mathcal{L}_b$	$\rightarrow$	Validation

Node splitting

- Conditional inference framework $\rightsquigarrow$ Generic algorithm

Stop criteria

- Generic algorithm $\rightsquigarrow$ No association between $\mathbf{X}$ and $\mathbf{Y}$
- Terminal nodes $\rightsquigarrow$ Minimum number of observations n_{min}

Conditional inference trees (II/III)
Generic node splitting algorithm

Key idea

- Variable selection, node splitting $\rightsquigarrow$ Separation
- Node representation $\rightsquigarrow$ Case weights $\mathbf{w}_m \in \{0, 1\}^N$

Selection step

- $H_0 : \mathcal{D}(\mathbf{Y}|\mathbf{X}, \mathbf{w}_m) = \mathcal{D}(\mathbf{Y}|\mathbf{w}_m)$ $\rightsquigarrow$ Test
- No rejection $\rightsquigarrow$ Stop
- Rejection $\rightsquigarrow$ Find $\mathbf{X}_{j*}$ with strongest association to $\mathbf{Y}$

Split step

- Take $\mathbf{X}_{j*}$ $\rightsquigarrow$ Find optimal binary split
- Daughter nodes $\rightsquigarrow$ Find case weights $\mathbf{w}_m^{\text{left}}$ and $\mathbf{w}_m^{\text{right}}$

Conditional inference trees (III/III)
Variable importance

Conditional permutation scheme

- Permutation of $\mathbf{X}_j$'s values in $\mathcal{O}_b \rightsquigarrow$ Mimic absence of $\mathbf{X}_j$
- Grid $\rightsquigarrow$ Cut-points of $\mathbf{X}_j^c = (\mathbf{X}_1 \dots \mathbf{X}_{j-1} \mathbf{X}_{j+1} \dots \mathbf{X}_p)$ in $\mathcal{T}_b$

Permutation-based importance measure

$$\text{VI}\left(\mathbf{X}_j\right) = \frac{1}{B} \sum_{b=1}^{B} \sum_{i \in \mathcal{O}_b} \left[\frac{\mathcal{I}\{y_i \neq \hat{y}_i(\mathcal{T}_b, \mathbf{X}_{\pi(j)|\mathbf{X}_j^c})\}}{|\mathcal{O}_b|} - \frac{\mathcal{I}\{y_i \neq \hat{y}_i(\mathcal{T}_b, \mathbf{X}_j)\}}{|\mathcal{O}_b|} \right]$$

Interpretation

- Prediction accuracy of $\mathbf{X}_j \rightsquigarrow$ Mean decrease

What is next?

Preliminaries

ARIMA model

$$\phi_p(B)\Phi_P\left(B^\tau\right)(1-B)^d\left(1-B^\tau\right)^D\{x_t\}=\theta_q(B)\Theta_Q\left(B^\tau\right)\{\varepsilon_t\}$$

Short form notation

- $(pdq)(PDQ)$

Classification rule

$(PDQ)=(000)$	$\rightarrow$	$\{x_t\}$	is non-seasonal (N-S)
$(PDQ)\neq(000)$	$\rightarrow$	$\{x_t\}$	is seasonal (S)

Candidate tests (I/III)
JD+ tests

Name	Variable checked for significance	Short
Modified QS test	Autocorrelations at seasonal lags	QS
Friedman test	ANOVA with repeated measures on intra-year ranks	FT
Kruskal-Wallis test	ANOVA without repeated measures on overall ranks	KW
Test for spectral peaks [1)]	Tukey and AR(30) spectra at seasonal frequencies	SP
Periodogram test	Weighted sum of periodogram at seasonal frequencies	PD
F-test on seasonal dummies	Effects of seasonal dummies in the "$(pdq)(000)$ + mean + seasonal dummies" model	SD

1 This test is not considered as a candidate test.

Candidate tests (II/III)
Medical (MD) and unit root (UR) tests

Name	Variable checked for significance	Short
Welch test	ANOVA with repeated measures on time series for heteroskedastic data	WE
Welch-Kruskal-Wallis test	ANOVA with repeated measures on overall ranks for heteroskedastic data	WEKW
Edwards test	Efficient score vector at seasonal harmonics with square root transformation	ED
Roger test	Efficient score vector at seasonal harmonics without square root transformation	RO
OCSB test [1)]	Effect of seasonal unit root in the "Mean + unit root + seasonal unit root + augmentations" model	OCSB

1 Three variants of this test are considered: OCSB1 uses augmentation lags $\{1, 2, 3\}$, OCSB2 uses augmentation lags $\{1, 2, 3, 12, 13\}$ and OCSB3 is OCSB2 with the disturbances being a MA(1) process instead of white noise.

Candidate tests (III/III)
ARIMA residual-based variants

Non-seasonal ARIMA model

$$\phi_p(B)(1-B)^d\{x_t\} = \mu + \theta_q(B)\{\varepsilon_t\}$$

Model identification

- Method $\rightsquigarrow$ Hyndman & Khandakar (2008)
- Restrictions $\rightsquigarrow$ $p \leqslant 3$, $q \leqslant 3$

Residual-based tests

- FT, KW, QS, RO, WE, WEKW $\rightsquigarrow$ Application to ARIMA residuals
- Short form $\rightsquigarrow$ Suffix "-R"

Overall test

Goal

- Model parsimony, prediction accuracy $\rightsquigarrow$ Potential conflict
- Interpretable results, low misclassification rates $\rightsquigarrow$ Balance

Step 1

- Most informative tests (MIT) $\rightsquigarrow$ Identification
- Conditional random forests (CRF) $\rightsquigarrow$ Recursive feature elimination (RFE)

Step 2

- Classification rule $\rightsquigarrow$ Derivation
- Combination of MIT $\rightsquigarrow$ Conditional inference tree

Step 1: identification of most informative tests (I/III)

RFE algorithm

Step (a): initialise

$\mathcal{L}$	$=$	$(\mathbf{XY})$		
$\mathbf{X}$	$=$	p-values of candidate tests	$\rightarrow$	Entire set
$\mathbf{Y}$	$=$	Seasonality dummy	$\rightarrow$	Simulated ARIMA models

Step (b): grow multiple CRFs

$\mathcal{L}$	$\rightarrow$	Independent samples $\mathcal{L}^{(i)}$, Validation (VAL) data $\mathcal{V}^{(i)} = \mathcal{L} \backslash \mathcal{L}^{(i)}$	$\rightarrow$	$i \in \{1, \ldots, L\}$
$\mathcal{L}^{(i)}$	$\rightarrow$	Bootstrap samples $\mathcal{L}_b^{(i)}$, OOB data $\mathcal{O}_b^{(i)} = \mathcal{L}^{(i)} \backslash \mathcal{L}_b^{(i)}$	$\rightarrow$	$b \in \{1, \ldots, B\}$
$\mathcal{L}_b^{(i)}$	$\rightarrow$	Conditional inference tree $\mathcal{T}_b^{(i)}$		

Step 1: identification of most informative tests (II/III)
RFE algorithm

Step (c): aggregate information

$$\bigcup_b \mathcal{O}_b^{(i)} \rightarrow \text{Importance } \mathrm{VI}^{(i)}(\mathbf{X}_j),$$
$$\mathrm{VI}(\mathbf{X}_j) = L^{-1} \sum_{i=1}^{L} \mathrm{VI}^{(i)}(\mathbf{X}_j) \rightarrow j \in \text{candidate tests}$$
$$\mathcal{V}^{(i)} \rightarrow \text{Overall misclassification rate}$$
$$\mathrm{MR}^{(i)} = \mathrm{MR}\left(\bigcup_b \mathcal{T}_b^{(i)}\right) \rightarrow \text{Mean/Median, SD}$$

Step (d): eliminate least important test

$$\mathbf{X} \rightarrow \mathbf{X}_* = \arg\min_j \mathrm{VI}(\mathbf{X}_j) \rightarrow \mathcal{L} = (\mathbf{X}_*^c \mathbf{Y})$$

Step 1: identification of most informative tests (III/III)
RFE algorithm

Step (e): repeat steps (b) to (d)

Stop criterion $\rightarrow$ $|\mathbf{X}| = p_{\text{min}}$ $\rightarrow$ RFE path

Step (f): select most informative tests

$$\text{RFE path} \rightarrow \text{Mean/Median}\left(\text{MR}^{(1)}, \ldots, \text{MR}^{(L)}\right), \text{SD}\left(\text{MR}^{(1)}, \ldots, \text{MR}^{(L)}\right), |\mathbf{X}| \rightarrow \mathbf{X}_{\text{MIT}}$$

Step 2: derivation of classification rule

Step (a): grow single decision tree

$$\mathcal{L} = (\mathbf{X}_{\text{MIT}}\,\mathbf{Y}) \rightarrow \text{Sample } \mathcal{L}^\star, \quad \text{Validation data } \mathcal{V}^\star = \mathcal{L} \setminus \mathcal{L}^\star$$

$$\mathcal{L}^\star \rightarrow \text{Conditional inference tree } \mathcal{T}^\star \rightarrow \text{Rule}$$

Step (b): prune single decision tree (optional)

$$\mathcal{T}^\star \rightarrow \text{Potential redundancies (depending on } |\mathbf{X}_{\text{MIT}}|, n_{\text{min}}, \ldots), \quad \text{Pruned conditional inference tree} \rightarrow \text{Simplified but equivalent rule}$$

What is next?

Model identification

Goal

- Construction of overall test $\rightsquigarrow$ "Representative" ARIMA models ($\tau = 12$, model shares, distributional properties)
- Source $\rightsquigarrow$ BBk macroeconomic time series database

Samples

- Monthly seasonal adjustment $\rightsquigarrow$ $3{,}300$ time series
- No seasonal adjustment $\rightsquigarrow$ $10{,}600$ time series

ARIMA identification

- JD+ $\rightsquigarrow$ regARIMA & TRAMO automatic routines
- Model shares p_{mk} $\rightsquigarrow$ Averaged share in class $k \in$ {N-S, S}

Model simulation

Weights

$$w_{mk} = \tilde{p}_{mk} \left(\sum_j \tilde{p}_{jk} \right)^{-1} \quad \text{with} \quad \tilde{p}_{mk} = p_{mk} \cdot \mathcal{I}\{p_{mk} \geqslant 0.01\}$$

Algorithm (Ollech & Webel, 2017)

- NORTA $\rightsquigarrow$ Logspline density estimation
- ARMA parameters $\rightsquigarrow$ Mimic multivariate distribution

Representative outcome

- $100{,}000 \cdot w_{mk}$ models $\rightsquigarrow$ 5, 10, 20 years
- Total $\rightsquigarrow$ $600{,}000$ ARIMA models

Construction of overall test

Candidate tests

- Three branches $\rightsquigarrow$ 18 tests
- Stationarity assumption $\rightsquigarrow$ Differencing (if necessary)

Step 1: RFE path

- Steps (b) to (d) $\rightsquigarrow$ $L = 50$ (size $\in [800; 8{,}000]$), $B = 100$, $n_{\text{min}} = 1$
- Step (c) $\rightsquigarrow$ VAL data $=$ Samples from $\mathcal{V}^{(i)}$ (size $= 50{,}000$)
- Step (e) $\rightsquigarrow$ $p_{\text{min}} = 1$

Step 2: classification rule

- $\mathcal{L}^{\star}$ $\rightsquigarrow$ Size $= 50{,}000$
- $\mathcal{T}^{\star}$ $\rightsquigarrow$ $n_{\text{min}} = 250$

What is next?

- Motivation
- Random forests and conditional inference trees
- Candidate and overall seasonality tests
- Simulation algorithm and design
- **Results**
- Summary
- References

Model identification
Top 10 ARIMA models

Simulation weights

As a percentage

Rank	Non-seasonal (N-S) class: Model m	Non-seasonal (N-S) class: Weight w_{mk}	Seasonal (S) class: Model m	Seasonal (S) class: Weight w_{mk}
1	(011)(000)	22.8	(011)(011)	47.5
2	(311)(000)	11.0	(010)(011)	8.3
3	(110)(000)	10.3	(311)(011)	7.7
4	(100)(000)	8.5	(210)(011)	6.6
5	(211)(000)	5.4	(110)(011)	6.4
6	(001)(000)	4.7	(211)(011)	4.4
7	(010)(000)	4.2	(012)(011)	4.0
8	(111)(000)	4.0	(111)(011)	3.6
9	(012)(000)	3.6	(011)(111)	2.5
10	(210)(000)	3.6	(010)(100)	2.0

Misclassification rates (I/III)

As a percentage, $\alpha = 0.01$

Classifier		Simulated ARIMA series All lengths N-S	All lengths S	5-year N-S	5-year S	10-year N-S	10-year S	20-year N-S	20-year S
CRF*	OOB	0.6	1.9	0.6	1.9	0.5	2.0	0.7	1.9
	VAL	0.6	1.9	0.6	2.0	0.5	1.8	0.6	1.8
JD+ tests	QS	4.9	1.5	2.5	1.7	5.0	1.4	7.1	1.3
	QS-R	0.3	8.4	0.2	7.5	0.2	7.8	0.6	9.8
	FT	2.1	2.1	1.5	2.2	2.3	1.9	2.4	2.1
	FT-R	0.8	2.1	0.4	2.3	0.8	2.0	1.4	2.1
	KW	2.4	3.8	1.9	3.9	2.6	3.7	2.7	3.8
	KW-R	0.7	2.1	0.3	2.2	0.7	2.0	1.1	2.2
	PD	3.2	3.6	3.2	3.4	3.3	3.6	3.2	3.9
	SD	4.0	2.7	4.4	2.5	4.1	2.7	3.7	2.9
MD tests	RO	11.8	93.9	10.1	93.7	11.6	94.0	13.7	94.0
	RO-R	16.7	78.7	15.6	82.3	17.2	77.6	17.2	76.4
	ED	4.2	99.2	3.3	99.1	4.1	99.3	5.2	99.4
	WE	3.5	3.8	4.3	3.7	3.2	3.7	3.0	4.0
	WE-R	1.7	2.1	2.7	2.0	1.5	2.1	1.1	2.4
	WEKW	5.6	3.4	9.0	3.1	4.5	3.4	3.4	3.7
	WEKW-R	4.1	1.8	7.6	1.5	2.9	1.8	1.7	2.0
UR tests	OCSB1	4.9	4.3	7.8	3.8	3.7	4.5	3.2	4.6
	OCSB2	3.3	4.9	8.4	3.1	1.0	5.0	0.6	6.6
	OCSB3	10.5	4.6	11.9	3.2	8.6	4.7	11.0	5.9

* Average misclassification rates over 50 training data sets at the inital RFE stage.

Misclassification rates (II/III)

As a percentage, $\alpha = 0.05$

Classifier		Simulated ARIMA series All lengths N-S	S	5-year N-S	S	10-year N-S	S	20-year N-S	S
CRF*	OOB	0.6	1.9	0.6	1.9	0.5	2.0	0.7	1.9
	VAL	0.6	1.9	0.6	2.0	0.5	1.8	0.6	1.8
JD+ tests	QS	7.4	1.2	4.9	1.4	7.5	1.1	9.8	1.2
	QS-R	1.1	7.0	0.9	6.1	0.9	6.4	1.4	8.3
	FT	6.6	1.6	5.7	1.6	7.0	1.5	7.2	1.7
	FT-R	4.2	1.6	3.2	1.6	4.2	1.5	5.2	1.7
	KW	6.9	3.2	6.2	3.1	7.1	3.1	7.4	3.3
	KW-R	4.0	1.6	3.1	1.6	4.1	1.5	4.7	1.8
	PD	8.1	3.2	8.3	2.8	8.2	3.1	8.0	3.5
	SD	9.1	2.2	9.7	2.0	9.2	2.2	8.5	2.5
MD tests	RO	15.1	91.0	13.5	90.7	14.9	91.1	16.8	91.3
	RO-R	19.9	71.7	18.8	74.9	20.6	70.6	20.4	69.8
	ED	4.9	99.1	3.9	98.9	4.8	99.1	6.1	99.3
	WE	8.9	3.3	10.8	3.0	8.2	3.2	7.7	3.6
	WE-R	6.4	1.7	8.6	1.4	5.7	1.6	4.9	2.0
	WEKW	12.0	2.9	17.5	2.7	10.1	2.9	8.3	3.2
	WEKW-R	10.4	1.4	16.7	1.1	8.3	1.4	6.1	1.6
UR tests	OCSB1	2.9	7.1	4.0	7.0	2.4	7.4	2.3	6.7
	OCSB2	0.8	10.0	1.7	6.9	0.4	9.6	0.4	13.4
	OCSB3	3.5	8.5	3.6	7.0	3.1	8.6	4.0	9.8

* Average misclassification rates over 50 training data sets at the inital RFE stage.

Misclassification rates (III/III)
Preliminary conclusion

Residual-based tests

- Lower $\rightsquigarrow$ Compared to standard tests (except for QS, RO)

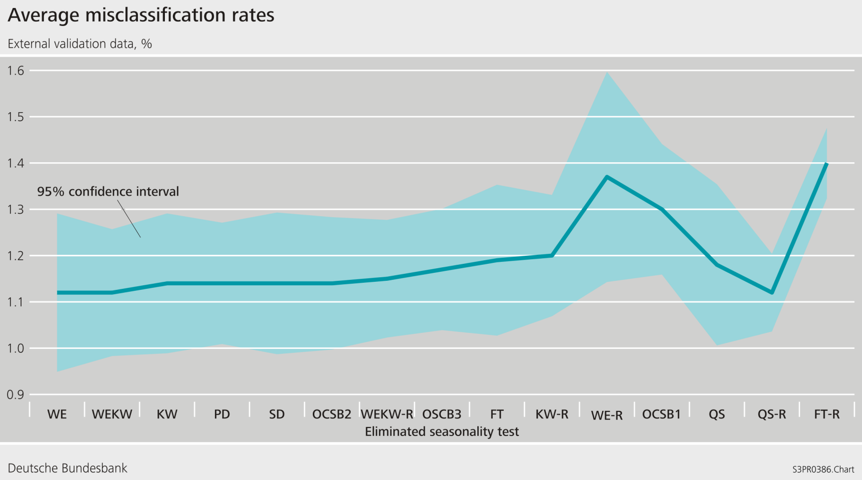
ED & RO tests

- Unacceptably high (especially for class S) $\rightsquigarrow$ No consideration for RFE path

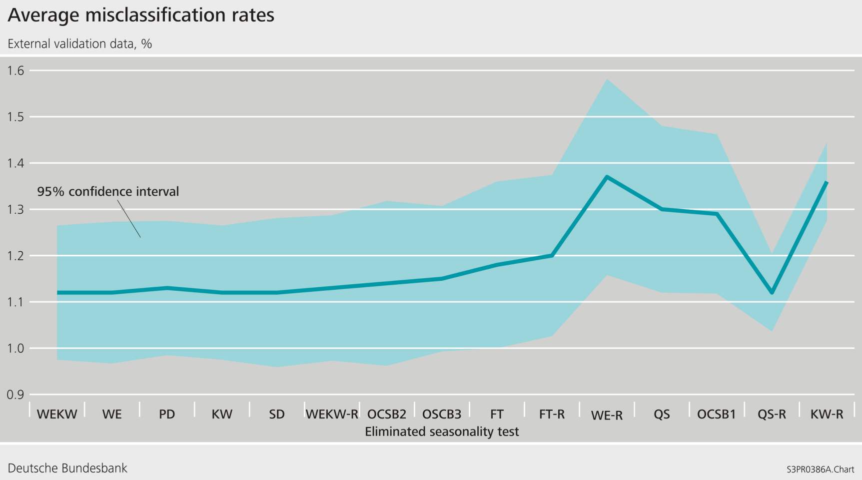
OCSB tests

- Lower for class S $\rightsquigarrow$ OCSB1
- Lower for class N-S (especially for longer series) $\rightsquigarrow$ OCSB2

RFE path (I/II)
Mean variable importance

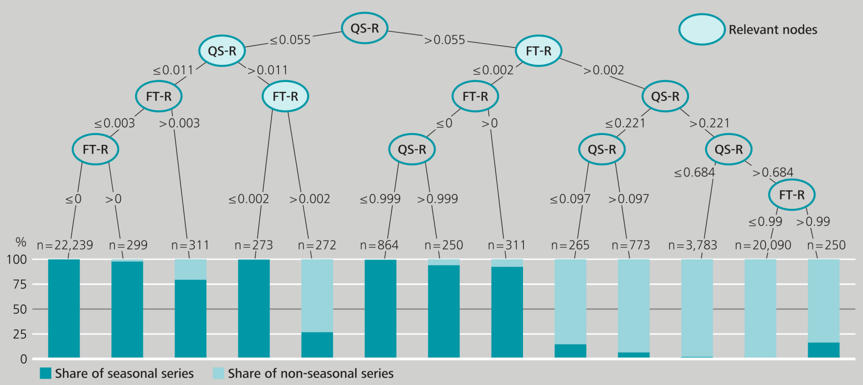


RFE path (II/II)
Median variable importance



Classification rule (I/III)
Conditional inference tree

Classification tree consisting of the QS-R and FT-R tests



Deutsche Bundesbank

Overall seasonality test

A **time series is** classified as **seasonal if**

1. **p-value $\leqslant$ 0.011 for** the **QS-R test** OR
2. **p-value $>$ 0.011 for** the **QS-R test** AND
   p-value $\leqslant$ 0.002 for the **FT-R test**.

Misclassification rates

As a percentage

	All lengths		5-year		10-year		20-year	
Data	N-S	S	N-S	S	N-S	S	N-S	S
Training set	0.54	1.56	0.18	2.20	0.36	1.47	1.09	1.02
Validation set	0.56	1.57	0.28	2.07	0.38	1.49	1.03	1.14

Classification rule (III/III)
Application

Are these economic time series seasonal?			
Test	Retail trade turnover: games and toys	HICP: tobacco	CPI: energy
QS	Yes	No	Perhaps
FT	Yes	No	Perhaps
KW	Yes	No	Yes
SP	Yes	No	No
PD	Yes	No	Yes
SD	Yes	No	Yes
Overall seasonality test			
QS-R	p-value $= 0.000$	p-value $= 0.969$	p-value $= 0.038$
FT-R	p-value $= 0.000$	p-value $= 0.740$	p-value $= 0.017$
Σ	Yes	No	No

What is next?

- Motivation
- Random forests and conditional inference trees
- Candidate and overall seasonality tests
- Simulation algorithm and design
- Results
- **Summary**
- References

In a nutshell

Overall seasonality test

- Most informative tests ⇝ Identification via RFE in CRF framework
- Classification rule ⇝ Derivation via pruned conditional tree

Large-scale simulation study

- ARIMA models ⇝ Representative of BBk database
- Most informative tests ⇝ ARIMA residual-based QS & FT tests
- Overall test ⇝ High precision, tractable rule

Future research

- Verification ⇝ Unbalanced training data (N-S vs S), barely seasonal data
- Seasonality tests ⇝ Very short series, $\tau \neq 12$

What is next?

- Motivation
- Random forests and conditional inference trees
- Candidate and overall seasonality tests
- Simulation algorithm and design
- Results
- Summary
- **References**

Random forests and conditional inference trees

- L Breiman (2001), Random Forests, Machine Learning 45 (1), 5-32.
- T Hothorn, K Hornik & A Zeileis (2006), Unbiased Recursive Partitioning: A Conditional Inference Framework, Journal of Computational and Graphical Statistics 15 (3), 651-674.
- H Strasser & C Weber (1999), On the Asymptotic Theory of Permutation Statistics, Mathematical Methods of Statistics 8, 220-250.
- C Strobl, A L Boulesteix, T Kneib, T Augustin & A Zeileis (2008), Conditional Variable Importance for Random Forests, BMC Bioinformatics 9, Article 307.

Simulation and automatic detection algorithms

- M C Cario & B L Nelson (1997), Modeling and Generating Random Vectors with Arbitrary Marginal Distributions and Correlation Matrix, Technical report, Northwestern University.
- R J Hyndman & Y Khandakar (2008), Automatic Time Series Forecasting: The forecast Package for R, Journal of Statistical Software 27 (3), 1-22.
- D Ollech & K Webel (2017), Assessing the Informational Content of Seasonality Tests by Random Forests of Conditional Inference Trees, JSM Proceedings, 1889-1900.
- C J Stone, M H Hansen, C Kooperberg & Y K Truong (1997), Polynomial Splines and their Tensor Products in Extended Linear Modeling, The Annals of Statistics 25 (4), 1371-1470.