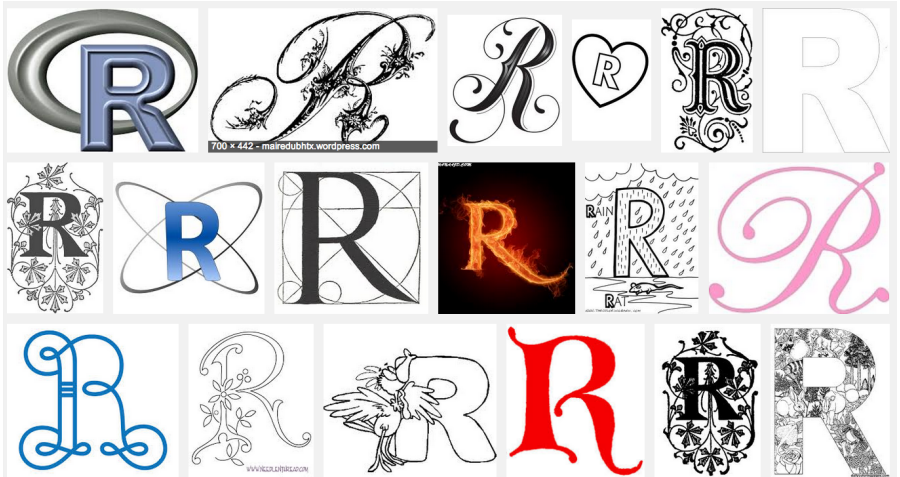


Matthias Templ,
Alexander Kowarik
Statistik Austria, Technische
Universität Wien
Bernkassel, 7. April 2014

Forschungsbereiche in der Offiziellen Statistik und Stichprobentheorie

- ▶ Fehlende Werte und Imputation
- ▶ (Stichprobenziehung und Kalibrierung, Matching)
- ▶ Robuste Schätzung von Indikatoren
- ▶ Saisonale Zeitreihenanalyse
- ▶ Statistische Geheimhaltung
- ▶ Mikrosimulationsstudien
- ▶ Visualisierung
- ▶ Kompositionsdatenanalyse
- ▶ Blended Learning

und zentral ist dabei immer ...





(visualization and imputation of missing values)

- ▶ Visualisierung der Struktur von fehlenden Werten
- ▶ Imputationsmethoden (hot-deck, kNN, robuster EM)
- ▶ Imputation von survey-Objekten
- ▶ Mehr als 150 Downloads pro Woche
- ▶ + Point and Click Graphical User Interface
- ▶ → Vortrag Morgen

- ▶ Praktische Simulationen: hilft Mikrozensus zur Verbesserung der Schätzung von EU-SILC Armutsindikatoren?
- ▶ Vergleich von Matching, Bootstrap und SAE Methoden
- ▶ MSE: bringt eigentlich nichts

- ▶ (Klassische und) Robuste semi-parametrische Schätzung von Indikatoren aus komplexen Stichproben.
- ▶ Kalibrierter Bootstrap
- ▶ Spezielle Klassenstruktur zur einfachen Handhabung von Indikatoren
- ▶ Journal of Statistical Software:
vignette('laeken-intro')

Domain: Burgenland, female older than 65.

(Loading movieginireal.mpg)

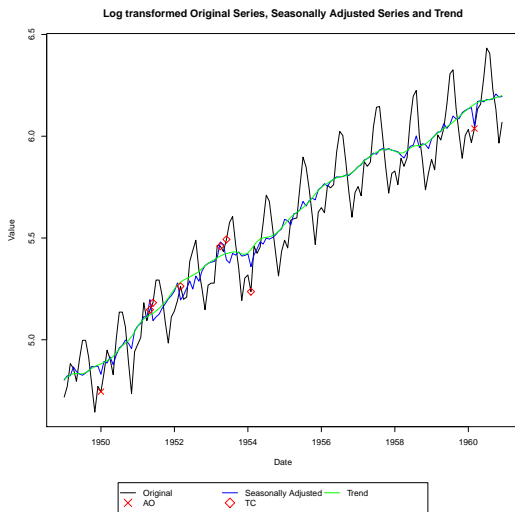
Domain: Salzburg.

(Loading movieQSRrealPDCout.mpg)

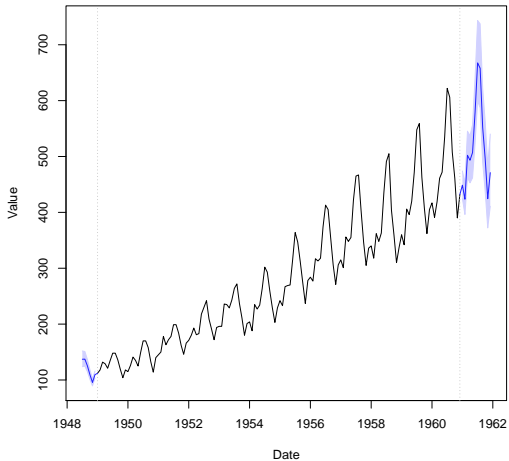


- ▶ X12ARIMA vom US Census Bureau ist Standardsoftware aber mühsam
- ▶ Output von X12ARIMA wird extrahiert
- ▶ batch-mode für multiple Zeitreihen
- ▶ Parameter können einfach verändert werden
- ▶ Interaktivität
- ▶ Grafiken (Ausreisser, Trend, Saisonfaktoren, ACF, ...)
- ▶ Journal of Statistical Software (2014)

x12: Time Series Plot with Outliers



Time Series with Back- and Forecasts



R Rgui.exe

X12 Export

timeseries

Series_1
Series_2
Series_3
Series_4

Span
 Start
Year: Period:
 End
Year: Period:

Modelspan
 Start
Year: Period:
 End
Year: Period:

Transform: auto

Arima
 Arima: 1 1 0
 SARima: 1 1 0

Settings
 Automdl
 Acceptdefault
 Balanced
 Seats
 Estimate
 Estimate out of Sample
 Sliding Spans
 OnlyTd
 stshort
 x11appendfcst
 x11appendbcst
 x11excludefcst

History
revert to:
previous
Revert
Clean History
Program started!

Manual Outliers
Type Year period
Remove
Type:
Year:
Period:
Add
Add by Click

Plot
 Seasonally Adjusted
 Trend
 Log-Transformation
 Show CI
 Original Points
 Show Allout
 Show Out
year:
period:
spectral
 sa
 original

autocorrelations of the residuals spectral seasonal factors plot summary text summary table

	Series_1	Series_2
X11 Regression	FALSE	FALSE
Model Definition	ARIMA Model: (0 1 1)(0 1 1) (Automatic Model Choice)	ARIMA Model: (0,1,1)(0,1,1)
Transformation	Automatic selection : Log(y)	Automatic selection : Log(y)
Regression Model	none	none
Outlier detection performed	FALSE	FALSE
Identifiable Seasonality	yes	yes
Seasonal Peaks	rsd	rsd
Trading Day Peaks	sa irr	sa irr
Q Statistic	0.26	0.26
Nr of M stats outside limits	0	0
SA decomposition	multiplicative	multiplicative
Seasonal moving average	3x3 (Based on msr size)	3x3 (Based on msr size)
Henderson filter	9-term	9-term
OLD OUTPUT 1	Run 1	NA
X11 Regression	FALSE	NA
Model Definition	ARIMA Model: (1,1,0)(1,1,0)	NA
Transformation	Automatic selection : Log(y)	NA
Regression Model	none	NA
Outlier detection performed	FALSE	NA
Identifiable Seasonality	yes	NA
Seasonal Peaks	none	NA
Trading Day Peaks	sa irr	NA
Q Statistic	0.26	NA
Nr of M stats outside limits	0	NA
SA decomposition	multiplicative	NA

R Rgui.exe

X12 Export

timeseries: stshort

Series_1 x11appendcst

Series_2 x11appendbcst

Series_3 x11excludcst

Series_4 x11regress

Maxorder: 3 2

Maxdiff: 1 1

Regvariables:

Reguser:

Regfile: (None)

usertype:

centeruser:

RegFileStart

Year: Period:

seatsparameter:

sigalim:

critical

all

AO LS TC

specific

History

revert to:

previous

Revert

Clean History

regvariables changed:

Type Year period

AO 1955 10

TC 1954 7

Remove

Type:

Year:

Period:

Add

Add by Click

Plot

Seasonally Adjusted

Trend

Log-Transformation

Show CI

Original Points

Show Allout

Show Out

year:

period:

spectral

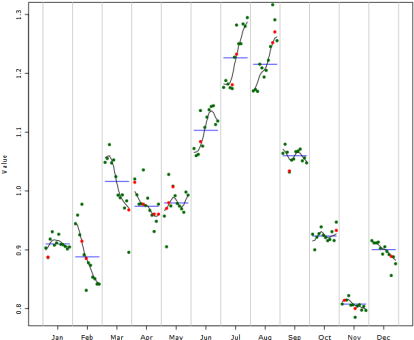
sa

original

irregular

autocorrelations of the residuals spectral seasonal factors plot summary text summary table

Seasonal Factors by period and SI Ratios



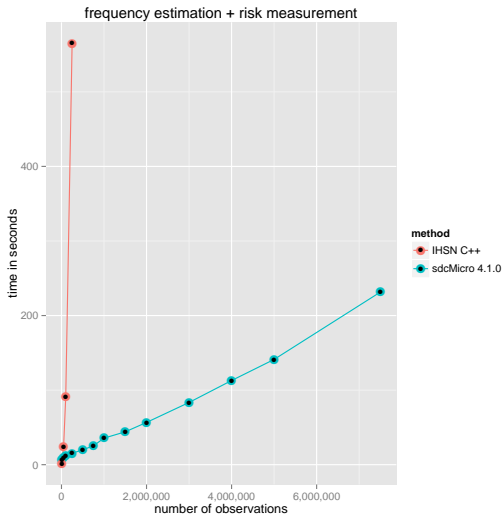
Value

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

Seasonal Factor
SI Ratio
Mean
Replaced SI Ratio

- ▶ Schutz vor Re-Identifizierung von mehrdimensionalen verlinkten Tabellen mit komplexer hierarchischer Struktur.
- ▶ Generierung eines SDC Problems
 - ▶ Spezifikation der Hierarchien
 - ▶ Daten: Mikrodaten oder Makrodaten ...
- ▶ Lösung des Sekundärsperproblems mit verschiedenen Methoden der linearen Programmierung

- ▶ Strukturierung der Daten in Software
- ▶ High-performance computing
- ▶ (fast) alle Methoden implementiert
- ▶ Guidelines und Software über Weltbank, OECD und IHSN verteilt.





- ▶ Synthetische Populationen aus Tabellen, Stichproben oder Census oder Kombinationen von diesen
- ▶ Kalibrierung der Populationen (0/1 Gewichte)
- ▶ Mikrosimulationsstudien
- ▶ wird uns die nächsten Wochen/Monate beschäftigen

League Table - Premier League 2011/12
















	Games	Performance	Wins	Draws	Losses	Points
Man Utd	38		23	11	4	80
Chelsea	38		21	8	9	71
Man City	38		21	8	9	71
Arsenal	38		19	11	8	68
Tottenham	38		16	14	8	62
Liverpool	38		17	7	14	58
Everton	38		13	15	10	54
Fulham	38		11	16	11	49
A. Villa	38		12	12	14	48
Sunderland	38		12	11	15	47
WBA	38		12	11	15	47
Newcastle	38		11	13	13	46
Stoke	38		13	7	18	46
Bolton	38		12	10	16	46
Blackburn	38		11	10	17	43
Wigan	38		9	15	14	42
Wolves	38		11	7	20	40
Birmingham	38		8	15	15	39
Blackpool	38		10	9	19	39
West Ham	38		7	12	19	33

Kleine Grafiken (sparks) in Text und grafischen Tabellen, wie sparklines

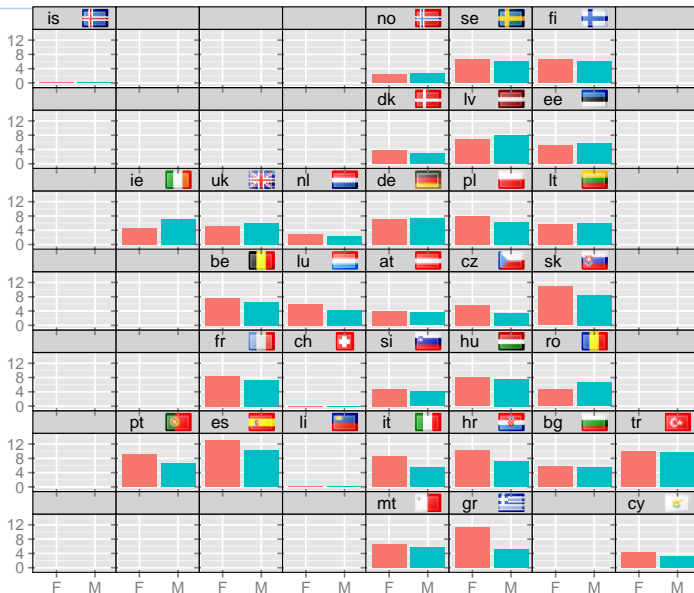
, Boxplots  und Balkendiagramme .

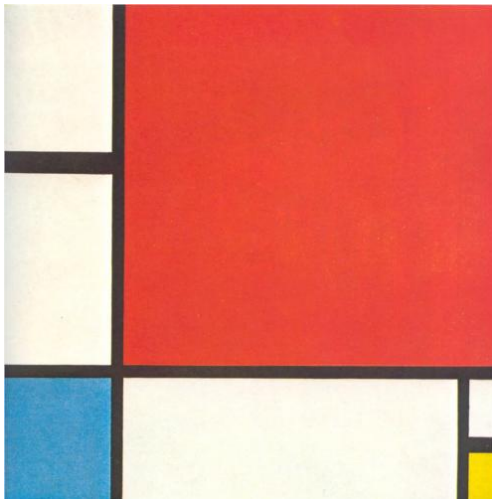
Finetuning is possible, e.g.   .

Zusätzliche Information

	2005-2010	2009 - 3	2009 - 4	2009 - 5	2009 - 6	2009 - 7	2009 - 8	2009 - 9	2009 - 10	2009 - 11	2009 - 12	2010 - 1	2010 - 2
BtisiF		103	101.7	100.5	101.6	102.6	102.2	103.5	105.2	104.8	104.3	104.2	101.6
BtisiE		102.1	99.8	99.2	100.2	101.8	100.9	102.6	104.8	103.8	103.6	103.6	101.4
EU_BCDF		102.7	101.1	99.8	100.9	102	101.9	102.4	104.5	104	103.9	105.1	101.4
EU_Industrie		101.2	98.8	98.4	99.1	100.9	100.7	100.8	104.1	102.6	103.4	104	100.6
Vorleistungen		98.2	92.8	93.4	95.2	96.6	96	98.7	98.2	100.2	99.6	99.6	98.6
Energie		119.9	121.3	118.7	118.9	118.4	115.6	117.5	119.8	118.8	118	114.9	115.4
Investitionsgüter		98.6	97.7	96.9	98.1	97.8	98.5	99.2	99.7	104.5	103	97.3	99.3
Konsumgüter		105.7	104.9	104	103.3	103.4	103.4	104.4	105.1	104.1	105.2	104.4	105.2
Lang_Konsumg		122.9	117.7	121.7	107.7	109.4	104.1	103.1	105.3	108.6	110.7	109.9	107.4
Kurz_Konsumg		103.1	103.1	101.5	102.6	102.1	103.3	104.3	104.5	103	103.7	103.2	103.8
Bergbau		103.3	105	108.3	114.5	115	111	110.1	108.7	111.9	102.8	98.6	105.7
HerstellungC		99.9	97.8	96.7	98	99.8	98.9	100.3	102.8	102.4	101.8	102	99.1
D		118.1	112.1	114.9	112	111.5	111.7	116.5	118.8	111.3	116.9	114.9	115.3
Wasser		122.4	132.4	130.6	132.9	135	134.6	133	128.1	125.8	124.3	128.8	134.2
Bau		107.3	109	105.9	107.6	106.7	107.6	107.7	107.4	109.3	107.3	106.2	102.7

Checkerplot





- ▶ Umgang mit strukturellen Nullern
- ▶ neue Theorie für Wahrscheinlichkeitstabellen



- ▶ Automatisierte Prüfungserstellung und Auswertung seit 2001 (Sweave)
- ▶ Blended Learning System TGUI (ab 2005)
- ▶ Feedback System, Server-Client Lösung
- ▶ Journal of Statistical Software
- ▶ <http://www.statistik.at/TGUIOnline>

- ▶ laeken
- ▶ (robCompositions)
- ▶ sparkTable
- ▶ sdcTable
- ▶ sdcMicro + sdcMicroGUI
- ▶ (TGUICore + TGUITeaching)
- ▶ VIM + VIMGUI
- ▶ x12

upcoming: CodaTable, simPopulation2, microSim

Der CRAN Task View **on Official Statistics and Survey Methodology** listed und beschreibt alle relevanten CRAN-Pakete mit folgenden Themen

- ▶ complex survey design;
- ▶ editing and visual inspection of microdata;
- ▶ imputation;
- ▶ statistical disclosure control;
- ▶ seasonal adjustment;
- ▶ statistical record matching;
- ▶ small area estimation;
- ▶ indices and indicators.

- ▶ Methodenentwicklung immer in Zusammenhang mit Anwendung und R Implementierung
- ▶ Zusammenarbeit sehr erwünscht
- ▶ Forschungsprojekte ansuchen