Outlier detection and seasonality breaks with JDemetra+ 3.0

using basic structural time series models

Jean Palate[1] Anna Smyk[2]

[1]National Bank of Belgium (NBB)

[2]Institut National de la Statistique et des Études Économiques (INSEE)

NTTS Conference, March 2021

Table of Contents

1 An alternative framework for Seasonal adjustment

2 Outlier detection

3 Seasonality breaks

4 Conclusion

Leading seasonal adjustment algorithms

- Seasonal adjustment methods recommended in the Eurostat guidelines: Tramo-Seats and X12-ARIMA (or X13)
- Both are "two-steps" methods:
  - Pre-processing step: removal of calendar effects and outliers by means of RegARIMA
  - Decomposition step: model-based approach (SEATS) or pre-defined filters (X11)

- Basic Structural Model (BSM) approach (state space form): alternative statistical framework for SA, considered by Durbin [1] more flexible, transparent and stable (?)
- One step method: integrated handling of pre-processing and decomposition (with various extensions)

An alternative framework for seasonal adjustment

- Key role of outlier detection
  - To avoid biased estimation of the parameters and spurious evolution of the hidden components
  - Most unstable part of the methods

- Aims of this work
  - Outliers detection in BSM. Extension of Grassi et al. [2], based on De Jong and Penzer [3]. See chapter 8 of the Eurostat handbook for seasonal adjustment.
  - New routines on BSM (including automatic outliers detection), available in an additional module of JDemetra+ 3.0 (version in development)

Basic Structural Model

The observed time series is expressed directly in terms of unobserved components: Let y_t denote a time series observed at t=0,1,...,n-1; the BSM is formulated as follows:

$$y_t = \mu_t + \gamma_t + X_t\beta + \epsilon_t, \quad t = 1, \dots, n,$$

where

$$\begin{aligned} \mu_t &= \text{local linear trend (or local level)} \\ \gamma_t &= \text{seasonal component of period } s \text{ (trig, HS, dummy...)} \\ X_t &= \text{explanatory variables or intervention effects} \\ \epsilon_t &= \text{noise} \end{aligned}$$

State space representation

- State:

$$\alpha_t = (\epsilon_t, \mu_t, \beta_t, \gamma_{i_t,t}, \ldots, \gamma_{j_t,t})'$$

- Measurement equation:

$$y_t = Z\alpha \quad \text{where} \quad Z = (1, 1, 0, 1, 0, \ldots, 0)$$

- Transition equation (defined by blocks):

$$\epsilon_{t+1} = 0\epsilon_t + \zeta_{\epsilon}{}_t \sim \mathrm{N}(0, \sigma_\epsilon^2)$$

$$\begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \zeta_{\mu t} \\ \zeta_{\beta t} \end{bmatrix} \quad \text{where} \quad \begin{cases} \zeta_{\mu}{}_t \sim \mathrm{N}(0, \sigma_\mu^2) \\ \zeta_{\beta}{}_t \sim \mathrm{N}(0, \sigma_\beta^2) \end{cases}$$

State space representation. Cont.

$$\begin{bmatrix} \gamma_{i_{t+1}t+1} \\ \vdots \\ \vdots \\ \gamma_{j_{t+1}t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \\ -1 & -1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \gamma_{i_t t} \\ \vdots \\ \vdots \\ \gamma_{j_t t} \end{bmatrix} + \zeta_{\gamma t}$$

where : $\zeta_{\gamma t} \sim \mathrm{N}(0_{s-1}, \sigma^2_\gamma \Omega)$

i_t is the (0-based) period of t, $j_t = i_t + s - 2 \mod s$

Remarks:

- Seasonal component: multivariate random walk; different disturbances covariance matrices Ω can be considered (corresponding to trigonometric, Harrison-Stevens, dummy... seasonal component)
- Noise in the state (no measurement error) $\rightarrow$ simpler derivations

Outlier detection algorithms

- Let D be any intervention variable corresponding to an outlier (additive outlier, level shift, seasonal outlier...).
- We consider the GLS model $y = D\beta + \mu$ where $\mu \sim N(0, \sigma^2\Omega)$.
- With $H_0 =$ no outlier, the usual statistic for the presence of an outlier is derived as follows:

$$\begin{aligned} S &= D'\Omega^{-1}D \quad , \quad s = D'\Omega^{-1}y \\ \hat{\beta} &= S^{-1}s \quad , \quad var(\hat{\beta}) = \sigma^2 S^{-1} \\ \tau_D^2 &= \hat{\sigma}^{-2}s'S^{-1}s \end{aligned} \tag{1}$$

- Extension to non stationary models: $\mu \sim$ any model (ARIMA, structural...)

Outlier detection Algorithms. Cont.

- Identification of a single outlier:
  - GLS solution: Compute τ^2 for each time point and each outlier (Tramo-Seats and X12-ARIMA approach)
  - BSM: shocks in the innovations correspond to the usual outliers $\rightarrow$ apply the Kalman smoother (KFS): (all τ^2 for any outlier are computed in one pass). Details in [3]
- Full algorithms
  - Forward iteration:
    - Outlier selection
      $$\begin{cases} \text{GLS: most significant outlier} \\ \text{BSM: most probable t for an outlier } \rightarrow \text{ outlier type} \end{cases}$$
    - (Partial) re-estimation of the model
  - Backward iteration: removal of non-significant outliers

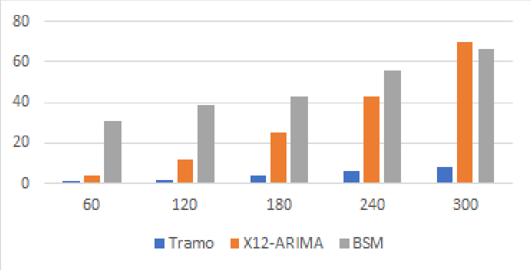
Comparison of the outlier detection algorithms

- Automatic outlier detection routines in ARIMA and BSM models provide comparable results
- Especially true when comparing BSM with airline models (similar statistical properties)
- Identical outliers (computed on 400 monthly series)
  - Tramo - BSM : 62 %
  - X13 - BSM : 67 %
  - Tramo - X13 : 73 %
- No significant differences between TRAMO, X13 and BSM concerning the stability of the detected outliers.

Comparison of the outlier detection algorithms (cont.)

- Speed performance (in JD+ 3.0)

Single outlier detection. Processing time by series length (fastest=1



- Tramo (RegArima approximations) much faster.
- BSM (KFS) slowest but rich output → alternative solutions

Seasonality break : definition

- Seasonality break = sudden change in the whole seasonal pattern (no additional assumption)
- KFS provides a straightforward way to test for the presence of a seasonality break in BSM:
  - We compute $\tau^2_{\gamma,t} = \max_{D_{\gamma,t}} \tau {D}^2_{\gamma,t}$ where $D_{\gamma,t}$ is generated by any shock in the seasonal block at t
  - See De Jong and Penzer (1998) [3] applied on the seasonal component
- Position of the seasonality break automatically identified ($\max_t \tau^2_{\gamma,t}$)

Seasonality break : modeling with state space forms

- Uni-variate form:

$$y_t = \mu_t + 1_{t<t_b}\gamma_{1_t} + 1_{t\geq t_b}\gamma_{2_t} + \epsilon_t$$

- Equivalent multi-variate form:

$$\begin{aligned} y_{1_t} &= \mu_t + \gamma_{1_t} + \epsilon_t \\ y_{2_t} &= \mu_t + \gamma_{2_t} + \epsilon_t \end{aligned}$$

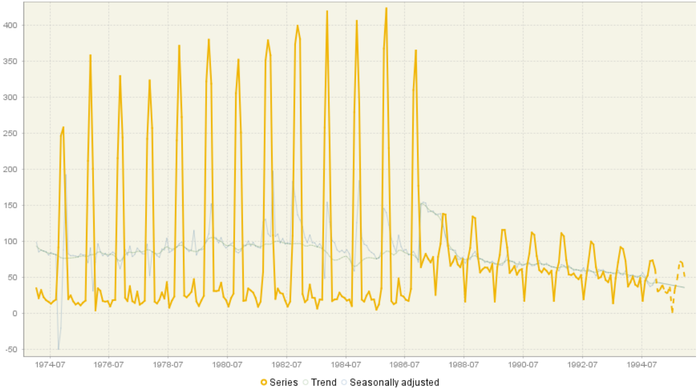
where

$$y_{1_t} = \begin{cases} y_t & \text{if} \quad t < t_b \\ \text{na} & \text{if} \quad t \geq t_b \end{cases} \qquad y_{2_t} = \begin{cases} \text{na} & \text{if} \quad t < t_b \\ y_t & \text{if} \quad t \geq t_b \end{cases}$$

- No reconciliation issue (common trend/seasonally adjusted series)

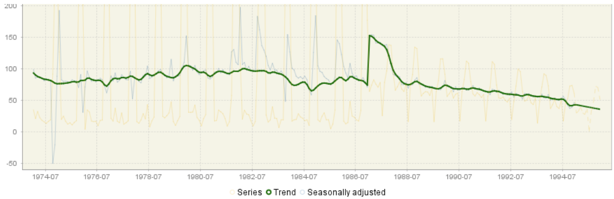
An extreme but real-life example of seasonality break (1)

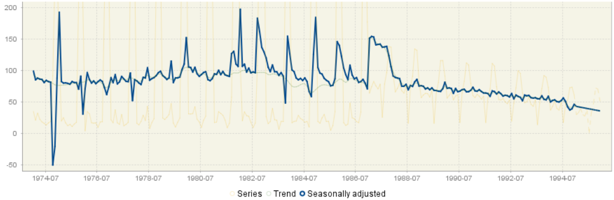
The sugar production in Belgium displays a significant seasonality break at the end of 1986, due to technical changes in the main factory Sugar production in Belgium



An extreme but real-life example of seasonality break (2)

Seasonal adjustment attempt with Tramo-seats (seasonal outliers allowed) : Trend and SA series





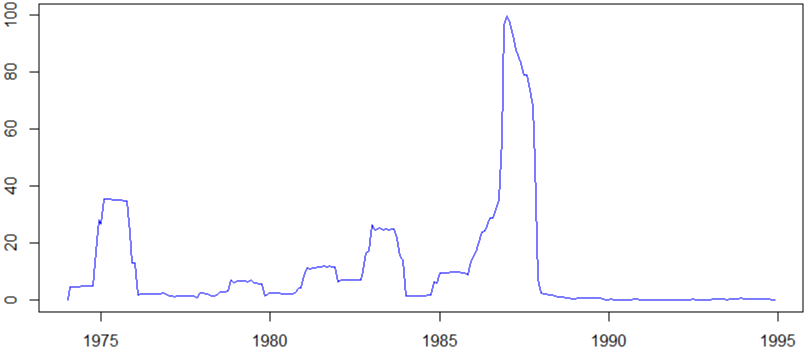
Regression model

Outliers

	Coefficients	T-Stat	P[|T| > t]
SO (11-1987)	-278,7919	-35,68	0,0000
SO (10-1987)	-215,1439	-29,14	0,0000
AO (1-1975)	107,8418	13,27	0,0000
SO (12-1987)	-153,6606	-19,72	0,0000
TC (10-1974)	-132,9650	-14,22	0,0000
SO (12-1983)	-143,1907	-19,72	0,0000
LS (1-1987)	80,3075	8,71	0,0000
TC (9-1982)	87,7097	10,32	0,0000
AO (12-1975)	-50,7437	-6,18	0,0000
SO (10-1979)	71,8113	11,06	0,0000
AO (1-1982)	90,7650	12,11	0,0000
SO (12-1981)	88,9883	12,39	0,0000
AO (12-1984)	71,9510	9,09	0,0000
TC (10-1985)	65,6748	7,56	0,0000
TC (11-1983)	94,6324	10,88	0,0000
AO (12-1979)	47,0435	6,18	0,0000
TC (12-1974)	76,8249	7,54	0,0000
TC (11-1984)	60,2597	6,88	0,0000
SO (7-1987)	-41,9155	-6,53	0,0000
TC (9-1981)	36,7089	4,32	0,0000
AO (11-1977)	-35,5716	-4,81	0,0000
TC (10-1983)	-39,6326	-4,67	0,0000
AO (5-1979)	29,0058	3,98	0,0001

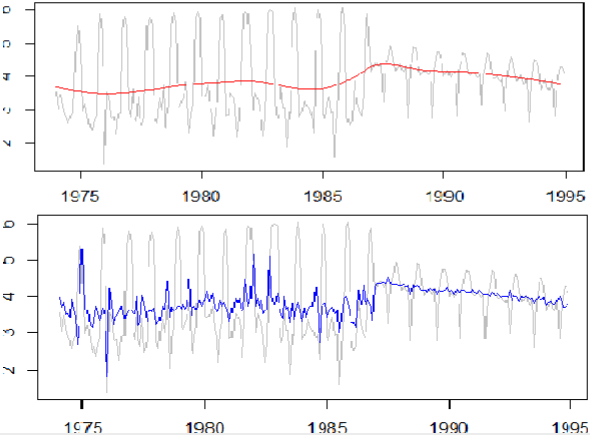
An extreme but real-life example of seasonality break (3)

Statistic τ_{γ}^{2} for the identification of seasonality breaks ($\sim \chi^2(11)$)



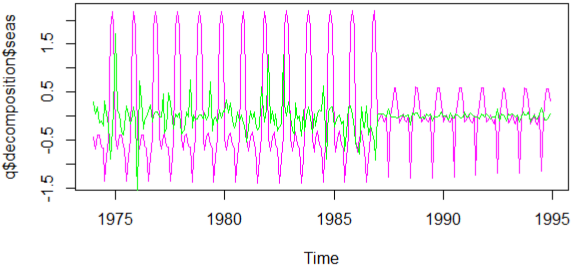
An extreme but real-life example of seasonality break (4)

Seasonal adjustment using multivariate BSM frame work:
Trend and SA series



An extreme but real-life example of seasonality break (5)

Seasonal adjustment using multivariate BSM framework: seasonal pattern and noise



Conclusion

- JDemetra+ 3.0, to be released in 2021, will offer additional seasonal adjustment methods based on BSM methods, including outlier detection and seasonality breaks
- Structural models (state space forms) offer an unified framework for pre-adjustment and decomposition of time series...
- ...which makes model-based advanced extensions such as time varying calendar effects or seasonal heteroskedasticity easier to implement

Thank you for your attention !

If you want to learn more about JDemetra+:

- https://github.com/jdemetra
- https://jdemetradocumentation.github.io/JDemetra-documentation

To download the R packages used for this presentation:

- https://github.com/palatej

Bibliography

James Durbin. "The foreman lecture: The state space approach to time series analysis and its potential for official statistics (with discussion)". In: *Australian & New Zealand Journal of Statistics* 42.1 (2000), pp. 1–23.

Stefano Grassi, Gian Luigi Mazzi, and Tommaso Proietti. "Automatic Outlier Detection for the Basic Structural Time Series Model". In: *Handbook on Seasonal Adjustment* (2018). URL: `ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001`.

Jeremy Penzer Piet de Jong. "Diagnosing Shocks in Time Series". In: *J. Am. Stat. Assoc.* 93.442 (June 1998), pp. 796–806.