Outlier detection and seasonality breaks with JDemetra+ 3.0

Keywords: *State Space Models, Seasonal adjustment, Outlier Detection, Seasonal Break*

1 Introduction

Outlier detection is a key step in any seasonal adjustment process. Identifying and removing inconsistencies from the series is necessary before estimating seasonal factors. Time series non corrected for outliers yield biased estimations of the parameters and spurious evolution of the hidden components (trend, seasonal), as underlined in part 2.1 of the Eurostat [1] guidelines for seasonal adjustment.

The seasonal adjustment methods currently recommended in the Eurostat guidelines are signal extraction ones, based on ARIMA models (Burman [2], Gomez and Maravall [3]) or semi-parametric ones, using predefined moving averages (Shiskin [4]). They are preceded by the detection and the removal of outliers and calendar effects by means of RegARIMA modelling. The seasonal adjustment process as a whole, pre-adjustment and decomposition, can be run with the TRAMO-SEATS algorithm(Gomez and Maravall [3]) or with the X13-ARIMA-SEATS algorithm (Monsell et al. [5]). Those two methods are available in the JDemetra+ [6] seasonal adjustment software, recommended by Eurostat and the European Central Bank.

An alternative statistical framework could be considered to achieve similar goals. The structural model approach proposed by Harvey [7], in which the observed time series is expressed directly in terms of unobserved components, is especially attractive. Durbin [8], in particular, discussed the use of such models in the context of official statistics, considering that they are more transparent, flexible and stable than ARIMA methods. The Eurostat guidelines on seasonal adjustment also mention that "unobserved component methods (. . .) based on state space models represent a reasonable alternative, provided they allow for a complete calendar and outlier treatment". Grassi et al. [9], whose work is featured in chapter 8 of the Handbook on Seasonal Adjustment, propose an automatic outlier detection procedure in the context of the Basic Structural Model (BSM). The statistical background for their algorithm is mainly provided by de Jong and Penzer [10].

Starting from their work, we have implemented in JDemetra+ 3.0 (version in development) an extension integrating automatic outlier detection and seasonal adjustment with BSM. It allows large scale processing and provides easy comparison with the standard outlier detection methods available in TRAMO-SEATS and in X13-ARIMA-SEATS.

A natural modification of the test proposed by de Jong and Penzer [10] leads to automatic detection of seasonality breaks, which correspond to a significant change in the whole seasonal pattern. Such changes are often more relevant than the usual seasonal outlier (Kaiser and Maravall [11]), too restrictive. Thanks to their flexibility, BSM models can be easily adapted to deal properly with series presenting seasonality breaks.

The first part of the paper focuses on automatic outlier detection in the BSM framework: our algorithm refines the solution proposed in the Handbook on seasonal adjustment. The second part explains the concept of seasonality break, describes the automatic test and presents the state space solution for estimating the corresponding model. We provide an example of real-world seasonality break for which the usual Arima based algorithms don't yield a satisfactory decomposition.

2 Methods

2.1 Automatic Outlier detection

The automatic outlier detection procedure implemented in JDemetra+ is designed to handle the basic structural model, which is described in many textbooks, such as Harvey [7] or Durbin and Koopman [12]. Alternative seasonal components to the usual trigonometric one can be specified, as detailed in Proietti [13]. Our algorithm focuses on additive outliers and on level shifts. (Seasonal outliers can optionally be considered) It aims at automatically determining the location, the type and the size of the outliers.

Roughly speaking, the Arima based automatic outlier detection procedure available in TRAMO-SEATS and in X13-ARIMA-SEATS and the one proposed in the Handbook for the BSM follow the same structure, with forward addition and backward deletion steps. They mainly differ on the method used in the forward steps to detect and measure the most significant outlier. The two ARIMA-based routines estimate a RegArima model for every possible outlier; the selection is defined by the highest T-Stat. As detailed in de Jong and Penzer [10], the Kalman smoother applied to the BSM can generate in a single run all the statics associated with the different outliers. We follow the suggestion of Grassi et al. [9] and select first the position, using a single criterion that corresponds to a joint test on the different outliers at a given time; the outlier type is selected in a second step. Such a solution is expected to be more robust than the T-Stat approach.

Our algorithm refines the one proposed in the chapter 8 of the Handbook on several technical details, among which the integration of additional regression variables such as calendar effects or diffuse initialization. Several options are available for model parameters' re-estimation between two forward or two backward steps: complete maximum likelihood re-estimation, partial re-estimation by means of one iteration of the score algorithm or no re-estimation. Score re-estimation in forward steps and no re-estimation in the backward steps seem to be the best trade-off between accuracy and performance. Finally, new default critical values have been computed by Monte-Carlo simulations, for different series lengths, different periodicities and different selection of outliers.

2.2 Seasonality break detection

The usual seasonal outlier patterns, for instance a shock in a specific period equally compensated by a small decrease in all the other periods, are, from our point of view, rather unrealistic. Global changes in the seasonal pattern as a whole - which we call seasonality breaks - are probably more relevant. They are indeed not uncommon in long series, which have often undergone a data collection or concept change and are in fact a superposition of two different data sets. Starting from the Kalman smoother output, we can, in a straightforward way, derive a test on the seasonal pattern of a BSM model as a whole. The aforementioned test is just a more targeted version of the single statistic discussed above. It also allows us to automatically date seasonality breaks.

In the presence of such breaks, the usual approach consists in splitting the series and applying separate seasonal adjustment procedures on the sub-series. However, the reconciliation of the resulting two seasonally adjusted and trend parts is often problematic. Multi-variate BSM models are particularly well-suited to deal with this kind of estimation problem. We consider that the series before and after the seasonality break are two separate series, only observed on a fraction of the time domain. Both series have a common trend (and perhaps irregular) but separate seasonal components. The state space framework of JDemetra+ can generate the usual decomposition and the seasonally adjusted series on the entire time span.

3 Results

3.1 Comparison of automatic outlier detection algorithms

Despite some defining differences, the automatic outlier detection routines in ARIMA and BSM models provide comparable results. That is especially true when airline models - which have rather similar statistical properties to those of BSM - are used. In our first tests on 100 real-world not too chaotic series, we found exactly the same set of outliers in more than 60% of cases (TRAMO vs BSM: 63%, X13 vs BSM: 68%, TRAMO vs X13: 73%). Moreover, the differences were marginal in most other cases. The greater robustness expected for the BSM routine has not been confirmed by actual tests. Using the stability of the detected outliers on variable time spans as an indicator, we have not observed significant differences between TRAMO, X13 and BSM. Further empirical investigations, building on those preliminary results, will be provided in the full paper.

3.2 An example of seasonality break correction

The sugar production in Belgium displays a significant seasonality break at the end of 1986, due to technical changes in the main factory. Figure 1 shows the decomposition of the series by means of a multi-variate BSM, after an automatic dating of the break.

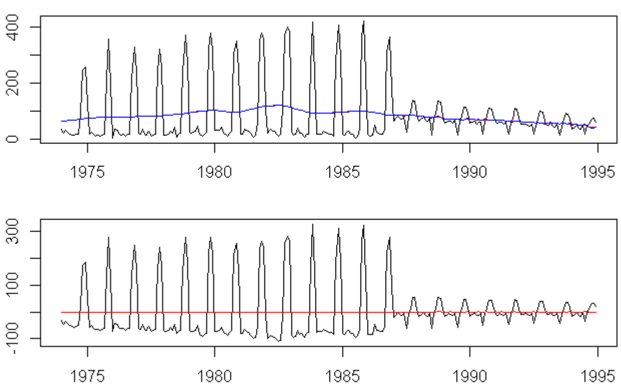

Figure 1: Top: raw, seasonally adjusted and trend. Bottom: seasonal factors and irregular

4 Conclusions

The 3.0 version of JDemetra+, to be released in 2021, will offer additional seasonal adjustment routines based on BSM methods, including outlier and seasonality break detection procedures. Basic structural models are less common than Arima models in seasonal adjustment algorithms and software. Nevertheless, they provide a unified framework for pre-adjustment and decomposition of time series, which also makes advanced extensions such as time varying calendar effects or seasonal heteroskedasticity easier to implement. These extensions will be presented in further papers.

References

[1] Eurostat. Ess guidelines on seasonal adjustment. Technical report, Eurostat Methodologies and Working Papers, European Commission, 2015. URL http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-15-001.

[2] J Peter Burman. Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society: Series A (General)*, 143(3):321–337, 1980.

[3] Victor Gomez and Agustin Maravall. Programs tramo (time series regression with arima noise, missing observations, and outliers) and seats (signal extraction in arima time series). instructions for the user. *Documento de Trabajo*, 9628:56, 1996.

[4] Julius Shiskin. *The X-11 variant of the census method II seasonal adjustment program*. Number 15. US Government Printing Office, 1965.

[5] Brian C Monsell et al. Us census bureau-x-13arima-seats seasonal adjustment program-home page. 2013.

[6] JDemetra+. Jdemetra+ software webpage. Technical report, 2020. URL https://github.com/jdemetra/jdemetra-app.

[7] Andrew C Harvey. Forecasting, structural time series models and the kalman filter. Technical report, 1989.

[8] James Durbin. The foreman lecture: The state space approach to time series analysis and its potential for official statistics (with discussion). *Australian & New Zealand Journal of Statistics*, 42(1):1–23, 2000.

[9] Stefano Grassi, Gian Luigi Mazzi, and Tommaso Proietti. Automatic outlier detection for the basic structural time series model. *Handbook on Seasonal Adjustment*, 2018. URL ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.

[10] Piet de Jong and Jeremy Penzer. Diagnosing shocks in time series. *J. Am. Stat. Assoc.*, 93(442):796–806, June 1998.

[11] Regina Kaiser and Agustín Maravall. Seasonal outliers in time series. *Documentos de trabajo/Banco de España, 9915*, 1999.

[12] James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*. Oxford university press, 2012.

[13] Tommaso Proietti. Comparing seasonal components for structural time series models. *International Journal of Forecasting*, 16(2):247–260, 2000.