Working Papers
09/2024

Improving quality in seasonal adjustment in Short-Term Statistics using JDemetra+ regressors and TEAM R-package

Félix Aparicio Pérez
José Fernando Arranz Arauzo
C. Gómez
E. Rosa-Perez
C. Sáez Calvo
L. Sanguiao Sande
María Teresa Vázquez Gutiérrez

First draft: July 2024

This draft: July 2024

Improving quality in seasonal adjustment in Short-Term Statistics using JDemetra+ regressors and TEAM R-package

Abstract

Short-term business statistics (STS) are the earliest statistics released to show emerging trends in the European economy. Monthly and quarterly STS provide data for the main economic sectors: industry, construction, trade and services, excluding financial and public services.

STS Regulation requires data that are calendar adjusted and calendar and seasonally adjusted, in addition to unadjusted data. Seasonal adjustment (SA) procedures eliminate the estimated seasonal and calendar effects from the original time series and obtain SA-estimates that are likely to reveal what is new in a time series.

JDemetra+ is the seasonal adjustment software officially used in the European Statistical System. Among other methods, it allows the use of a model-based TRAMO-SEATS approach for performing seasonal adjustment of time series. In this approach, a RegARIMA model is fitted to the series. It also offers the chance to calculate regression variables to model calendar effects, including trading day regressors that take into account the composition of the days of the month.

ARIMA models used to adjust STS time series play a very important role to obtain accurate adjusted data. But sometimes the work of updating the ARIMA models can become a burdensome task as the manual identification of a suitable model can become complex and time-consuming and automatic procedures can provide a model without taking into account some restrictions considered as essential for the domain expert.

Time-Series Exhaustive Automatic Modelling (TEAM) is an R package developed by Statistics Spain, and based in the JDemetra+ ecosystem, that can help in the process of updating ARIMA models by providing a list of models ordered according to a global score.

Methodologically, we can set a priori specifications (outliers, calendar regressors, maximum/minumum values of the ARIMA parameters), and the local scores by hierarchical levels can help us guarantee the quality. In this sense, one of the main advantages of TEAM is that the ARIMA models provided can be subject to some restrictions specified by the domain expert.

To carry out calendar adjustment at Statistics Spain we have been using customized working days regressors. However, for time series of specific activities the residual effects of trading days were not being completely removed. Several analyses have been undertaken and improvements have been achieved when using JDemetra+ regressors.

Due to the increase in quality, the JDemetra+ regressors and the use of the TEAM package to update the ARIMA models are included in the seasonal adjustment process as from January 2024 in STS at Statistics Spain.

Keywords

seasonal adjustment, JDemetra+, R, Time-Series Exhaustive Automatic Modelling

Authors and Affiliations

Félix Aparicio Pérez, S.G. for Methodology and Sampling Design

José Fernando Arranz Arauzo, S.G. for Information Technologies and Communications

C. Gómez, S.G. Short-Term Statistics

E. Rosa Pérez, S.G. Short-Term Statistics

C. Sáez Calvo, S.G. for Methodology and Sampling Design

L. Sanguiao Sande, S.G. for Methodology and Sampling Design

María Teresa Vázquez Gutiérrez, S.G. for Information Technologies and Communications

INE-Spain

1. Introduction

Short-term business statistics (STS) are the earliest statistics released to show emerging trends in the European economy. Monthly and quarterly STS provide data for the main economic sectors: industry, construction, trade and services, excluding financial and public services.

STS data are disseminated in form of indices together with annual and monthly rates. These rates can provide information about the change that has happened from one year to another, or from one month to another.

Time series data may sometimes exhibit strong periodic patterns. This is often referred to as the time series having a seasonal behaviour. This mostly occurs when data are taken in specific intervals (monthly, weekly, and so on). In these cases, the annual rate may not be a good indicator of the economic situation, and the same could happen with the monthly rates.

STS are provided on the basis of the European business statistics (EBS) regulation (Eurostat 2019, 2020), and, in order to disseminate comparable indices, the data requirements include data that are calendar adjusted and seasonally adjusted, in addition to unadjusted data.

The main objective of seasonal adjustment is to remove the seasonal fluctuations from the monthly (or quarterly) data series.

The calendar effects are those parts of the movements in the time series that are caused by different number of weekdays in calendar months. They arise as the number of occurrences of each day of the week in a month differs from year to year and from month to month. These differences cause regular effects in some series. Other calendar effects are moving holidays, which may fall in different months from year to year, Easter is the most important in Europe, and leap year effect because of an extra day inserted into February every four years.

Despite all the advantages of the calendar adjusted and seasonally adjusted data, we should also keep in mind that among others:

- The series that are adjusted for seasonal and calendar effects depend on the method used and on the assumptions of the chosen model, as well as on the software used in the process.
- An inappropriate or poor quality seasonal adjustment can generate erroneous results and false signals.
- The presence of residual seasonality or excessive smoothing negatively affects the interpretation of the seasonally adjusted data.

As it is mentioned in the second point, the quality of the adjustment is really important to disseminate accurate results. This quality should be measured not only for the statistical output, but also for the statistical processes, as we claim in this paper.

One of the European statistics principles in the European Statistics code of practice (Eurostat, 2017) is the comparability, between regions and countries, and over time. The seasonal and calendar adjustment is the way to assure this comparability over time.

Another principle related to the quality is the use of a sound methodology, and using a methodology created and shared by the countries in the European Statistical System (ESS) has many advantages as having a comparable methodology, saving cost and increasing the efficiency. All in one, increasing the quality.

In this paper we will present the main problems that Statistics Spain has faced related to the calendar and seasonal adjustment and how some of them have been solved with the use of JDemetra+ default regressors and the TEAM (Time-Series Exhaustive Automatic Modelling) package, an R package designed and developed by the methodology and the ICT units at Statistics Spain (Aparicio et al., 2024). We will also point out which principles of the European Statistics Code of Practice (ESCoP) are being fulfilled after the change.

2. Methodology

The seasonal and calendar adjustment provide an estimate of what is "new" in a series (change in the trend, the cycle and the irregular component). This part of the statistical process is really important, as seven of the 27 Principal European economic indicators (PEEI) are STS indicators, and the unit of measure of these PEEIs at Eurostat website is the seasonal and calendar effect adjusted index.

JDemetra+ (JD+) is an open-source software for seasonal adjustment developed by the National Bank of Belgium in cooperation with the Deutsche Bundesbank and Eurostat in accordance with the ESS guidelines on seasonal adjustment (Eurostat, 2015). It has been officially recommended by Eurostat to the ESS members since 2015 (Eurostat, 2018) and is being used in STS at Statistics Spain since 2018.

In addition, beyond seasonal adjustment, with JD+ it is also possible to calculate temporal disaggregation and benchmarking, now-casting and to build regressors to be used for calendar adjustment.

2.1 Calendar effects adjustment

Typical calendar effects include the different number of working days in each month, the different composition of the number of working days, the leap-year effect, and moving holidays.

The calendar effects adjustment has the purpose of obtaining series whose values do not depend on the length of the month, nor on the composition of the number of working days. It may be noted that the length and composition of each month has a seasonal part, so the working-day adjustment must only be associated with the non-seasonal part of the effect, which is the deviation of this number of days from its long-term monthly or quarterly average, calculated over a 28-year calendar.

Therefore, in order not to eliminate seasonality, it is necessary to calculate the regressors in deviations from the average for each month. The regressor for working days, thus built, is recalculated every year since the beginning of the series with the new averages of working and non-working days each month and the new proportion between working and non-working days. This regressor is called working day regressor (WD).

It is recommended to test whether there are differences between the different working days, mainly in long monthly series that measure the evolution of some variable or phenomenon in which, a priori, the behaviour may be different of each day of the week or for some of them.

If there are significant differences between each weekday, it is recommended to use a regressor for each day (from Monday to Saturday), called trading day regressors (TD).

Every four years, February has 29 days, which affects the comparability of the time series two ways: in the first place, the composition of the month varies (plus one weekday) and, in second place, the length of the month changes. The first effect is measured with the TD regressor; whereas the second effect must be quantified with the leap year regressor, whose value will be 0.75 in the months of February in leap years, -0.25 in the months of February in non-leap years and 0 in the remaining months.

The adjustment for moving holidays aims at eliminating from the series those values that are affected by events following a complex pattern over the years. The only moving holiday in Spain is Easter. This effect is also partially seasonal, as on average, it takes place more often in April than in March. Given that the seasonal part must be captured in the seasonal component, it must not also be eliminated with the Easter regressor (EA).

Finally, to calculate calendar regressors for calendar adjustment the use of national calendars is recommended at the Member State level (Eurostat, 2015).

Spanish NUTS2 regions have their own holiday calendar additional to the national bank holidays. This means that in some regions some specific days are non-working, while in other regions they are working days. This can affect the time series in a different way depending on the economic activity and the region, being necessary to calculate appropriate weights to have the regional calendar taken into account in the calendar adjustment.

Seasonal and calendar adjustment in STS at Statistics Spain is carried out since 2013 in Base 2010. At that moment, the software used was TRAMO-SEAT, and Statistics Spain designed and calculated customized WD and EA regressors. An advantage of these regressors was to account for bank holidays in each Spanish NUTS2 region. However, as WD regressor only differentiate between weekdays and weekend days, in certain time series of specific activities the residual effects of trading days were not completely removed.

Since Base 2015 the software used in STS is JDemetra+, and in Base 2021 the TD and EA regressors calculated by default by JDemetra+ are being used in the seasonal and calendar adjustment, but taking into account only the national bank holidays, not the regional ones.

To assess the improvement in the use of the JDemetra+ regressors as compared with the customized Statistics Spain regressors the time series that showed lower residual effects of trading days for turnover in Industry (ITI), in Retail Trade (RTI) and in Services (SSAI) and production in Industry (IPI) were selected. A comparison was made with the results obtained when the seasonal and calendar adjustment was carried out with the following regressors: JD+ TD with national bank holidays calendar & JD+ EA, JD+ TD with euro area calendar & JD+ EA, JD+ TD with national bank holidays calendar & INE EA, JD+ TD with euro area calendar & INE EA, INE WD & JD+ EA, INE WD & INE EA.

JDemetra+ assesses the results as: Good, Uncertain, Bad or Severe (JDemetra+ Reference Manual). In order to rank the analysis carried out, we punctuated good with 3 points, uncertain with 2 points, bad with 1 point and severe with 0 points. In the following table we can see the results for each statistical operation.

Table 1: Raking of regressors combination in STS indicators

SSAI		RTI		ITI		IPI	
JD+ TD (nat) & EA	2.84	JD+ TD (nat) & EA	3.00	JD+ TD(EU) & EA	2.87	JD+ TD(nat) & EA	2.38
JD+ TD (nat) & INE EA	2.80	JD+ TD(EU) & EA	3.00	JD+ TD(EU) & INE EA	2.83	JD+ TD (nat) & INE EA	2.34
JD+ TD(EU) & EA	2.67	JD+ TD (nat) & INE EA	2.97	JD+ TD (nat) & INE EA	2.27	JD+ TD(EU) & EA	2.32
JD+ TD(EU) & INE EA	2.65	JD+ TD(EU) & INE EA	2.81	JD+ TD (nat) & EA	2.19	JD+ TD (nat) & EA	2.25
INE WD & JD+ EA	1.82	INE WD & INE EA	1.04	INE WD & JD+ EA	1.43	INE WD & INE EA	1.78
INE WD & INE EA	1.69	INE WD & JD+ EA	0.90	INE WD & INE EA	1.17	INE WD & JD+ EA	1.75

Although the results vary from one STS indicator to another, the use of JDemetra+ TD regressor improves the accuracy of the seasonal adjustment.

2.2ARIMA model selection and outlier detection and correction

As we have already mentioned, JDemetra+ is the seasonal adjustment software officially used in the ESS. Among other methods, it allows the use of a model-based TRAMO-SEATS approach for performing seasonal adjustment of time series. In this approach, a RegARIMA model is fitted to the series.

An adequate model selection includes, among others, decision making regarding whether or not to apply the logarithmic transformation, the adequate differentiation orders in the regular and seasonal parts, the use of additive or multiplicative components or the statistical tests in order to assess the adequacy of the model.

ARIMA models used to adjust STS time series play a very important role to obtain accurate adjusted data. But sometimes the work of updating the ARIMA models can become a burdensome task as the manual identification of a suitable model can become complex and time-consuming and automatic procedures can provide a model without taking into account some restrictions considered as essential for the domain expert.

Outliers are abnormal or atypical values in the series. These may appear in different forms: as an impulse/additive outlier (AO), a transitory change (TC), a level shift (LS) or a change in the variation rate, depending on the underlying reason behind the outlier and its length.

These values must be detected and modelled before the seasonal and the calendar adjustment takes place, in order to avoid the estimates being distorted or biased.

According to the ESS guidelines on seasonal adjustment (Eurostat, 2015), the update of the ARIMA models should be undertaken at least once a year. At Statistics Spain (INE Spain, 2024) it was decided to identify and fix the model (including: outlier intervention variables and regressors for the correction for calendar effects) once per year.

However, in the last years, many events have taken place that have influenced the way the model selection and the outlier detection have been carried out at Statistics Spain.

In year 2020 the Covid pandemic affected all the economic sector for a longer or shorter period of time and in Spain many industries and businesses remained closed for weeks or even months. As a consequence, AOs were detected and corrected on a monthly basis all through the year. Once December 2020 data was disseminated, a thorough analysis took place and some of these outliers were changed to TCs or even LSs depending on the sector and the effect caused on the economy.

In 2021 economic recovery took place, but not at the same speed in the different sectors, and in Spain there were still some lockdowns in some regions. This situation demanded to keep the monthly AOs detection all the year round together with an additional analysis at the end of the year to remove some AOs or change them to LSs or TCs.

In March 2022 the Ukrainian war broke out, affecting some economic activities, and for that reason the monthly outlier detection and the annual analysis was maintained. In 2023 the exercise of detecting monthly outliers reflected that, although some AOs were still detected,

the number of cases was low and there was no need to carry on in 2024 with this monthly exercise.
So the analysis of the outliers during the period 2020-2023 was carried out on a monthly basis and revised annually, having all of them a clear interpretation according to the knowledge of the phenomenon by the expert. This in-depth exercise recommended that the outliers of this period should be kept fixed in the future.
Regarding the update of the ARIMA models, through the years 2021 to 2023, because of the unstable economic situation no model change was made, just the annual update of outliers. So there was a need to revise the ARIMA models to be used in year 2024, but taking into account some restrictions considered as essential by the domain expert.
TEAM is an R package developed by Statistics Spain, and based in the JDemetra+ ecosystem, that can help in the process of updating ARIMA models by providing a list of models ordered according to a global score. This package was designed and developed by the methodology and ICT units at Statistics Spain to meet the domain experts' needs.
Methodologically, we can set a priori specifications in TEAM (outliers, calendar regressors, maximum/minumum values of the ARIMA parameters), customized by the domain experts using the R script to set the specifications and call the TEAM package, and the local scores by hierarchical levels to be used to calculate the global score can help us guarantee the quality.
As the TEAM package is based in the JDemetra+ ecosystem, we use a Workspace as a base for the update of the ARIMA models. The initial workspace contains the outliers already fixed for the whole length of the time series. This saved a lot of time in the automatic model selection and avoided the inclusion of new outliers that could have no meaning for the domain experts.
As we have already mentioned in the previous section, the regressors used in 2024 were changed from the customized INE ones to those in JDemetra+, and it was carried out with the script using TEAM. There was no need to carry out any manual work, avoiding thus possible human errors.
Finally, TEAM makes it also possible to assess the adequacy of several ARIMA models (logarithmic transformation, orders in the regular and seasonal part, etc.) with the fixed outliers and the JDemetra+ regressors. In this sense, one of the main advantages of TEAM is that the ARIMA models provided can be subject to some restrictions specified by the domain expert, like fixing the differencing orders or setting the maximum or the minimum value of the parameters, or the outliers for some specific period different to the whole length of the time series.
This way we can obtain a workspace with as many models as we set in the R script ranked according to a global score based on weighting four components: model specification, signal

extraction, revision and residual seasonality. The local scores of these four components can be adjusted by the domain expert if he considers that some component should weight mores/less than other(s).
The expert can then choose the final model to be used in the annual seasonal adjustment process taking into account all the information included in the workspace and the files created during the process.

3. Conclusions and further work

As we have seen in this contribution, the use of the JDemetra+ TD regressors has improved the quality of the calendar adjustment in STS at Statistics Spain. We have implemented a sound methodology (ESCoP principle 7) as the formulas of the regressors have been defined by the European experts in the Centre of Excellence on Time Series Analysis and Seasonal Adjustment and we are using European guidelines, and good practices, while striving for innovation. Any improvement in the definition or calculation of these regressors will be automatically implemented. Statistics Spain is using appropriate statistical procedures (ESCoP principle 8) as the statistical processes are routinely monitored and revised as required not only by the national statistical system, but by the ESS. And finally Statistics Spain has improved the effectiveness (principle 10) with the implementation of standardized solutions.
If we consider the statistical output principles included in the ESCoP, the use of the JDemetra+ TD regressors has helped Statistics Spain to increase the accuracy (ESCoP principle 12) of the calendar adjustment and it has also improved the comparability (ESCoP principle 14) between countries, since Spain is now using the same regressors that are available to the other countries.
There is still further work to be developed related to the bank holidays in the Spanish NUTS-2 regions. So far, the calendar being used in JDemetra+ includes only nine bank holidays out of the 14 public holidays in Spain. The nine ones included are those common to all the NUTS-2 regions. There are another 3 days specific for each NUTS-2 region and another 2 specific for each Spanish municipality. So, we will have to analyse how to include the effect of the specific bank days on the calendar adjustment and its impact on the results as a possible improvement in the calendar adjustment.
Regarding the use of TEAM as the package to set the ARIMA model, it has proved the commitment to quality (ESCoP principle 4), as Statistics Spain identified a weakness and developed the tools to improve the process quality. We are now using a sound methodology (ESCoP principle 7) shared by different domains at Statistics Spain (STS, National Accounts) that provide feedback to try to improve it. Regarding the cost effectiveness (ESCoP principle

10) it has saved a lot of time, and the use of a standardised solution will help us in the future to make better decisions.

References (12pt. bold)

Aparicio Pérez, F., J. F. Arranz Arauzo, C. Sáez Calvo, L. Sanguiao Sande, and M. T. Vázquez Gutiérrez. Time-Series Exhaustive Automatic Modeling: a new methodology for model identification. European Conference on Quality in Official Statistics (Q2024), Estoril (Portugal) 5-7 June 2024

Eurostat (2015). ESS guidelines on seasonal adjustment. Available at: https://ec.europa.eu/eurostat/documents/3859598/6830795/KS-GQ-15-001-EN-N.pdf

Eurostat (2017). European Statistics Code of Practice. Available at: https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7?t=1528447068000

Eurostat (2018). Handbook on Seasonal Adjustment. Available at: https://ec.europa.eu/eurostat/documents/3859598/8939616/KS-GQ-18-001-EN-N.pdf

Eurostat (2019, 2020). Regulations Commission Implementing Regulation (EU) 2020/1197 laying down technical specifications and arrangements pursuant to Regulation (EU) 2019/2152 of the European Parliament and of the Council on European business statistics repealing 10 legal acts in the field of business statistics.

INE Spain (2024). INE standard on seasonal and calendar adjustment. Available at: https://ine.es/en/clasifi/estandar_efectos_estacionales_en.pdf

JDemetra+ Reference Manual. Available at: https://jdemetra-new-documentation.netlify.app/