Instituto Nacional de Estadística

Working Papers

07/2024

Time-Series Exhaustive Automatic Modeling: a new methodology for model identification

Félix Aparicio Pérez
José Fernando Arranz Arauzo
Carlos Sáez Calvo
Luis Sanguiao Sande
María Teresa Vázquez Gutiérrez

First draft: July 2024

This draft: July 2024

Time-Series Exhaustive Automatic Modeling: a new methodology for model identification

Abstract

Seasonal adjustment of time series plays a pivotal role in modern official statistics, ensuring accurate and reliable data analysis. However, due to resource constraints and time limitations, the models identified in an automatic way using the current software may not be optimal. This leads to a worse performance of seasonal adjustment, since these models must be maintained for a year.

We present a new R package, Time-Series Exhaustive Automatic Modeling (TEAM), which aims to automate and enhance the yearly model identification phase. The goal is to provide in an automated way a list of optimal models, where the optimality criteria can be specified by the users to meet their specific needs.

The methodology employed in TEAM is characterized by an exhaustive search and ranking of models. Initially, an exhaustive search of specifications is conducted for each time series, testing all possibilities for parameters such as data transformations (logarithms or levels), the order of the ARIMA model, inclusion of outliers, and calendar regressors. Subsequently, each specification is processed using the JDemetra+ software in a parallelized way, yielding diagnostic information to construct five indicators assessing the model's performance across distinct areas.

The five indicators and their respective areas of evaluation are as follows:

1. Model Diagnostics: Measures the model adequacy by using the statistical tests on the residuals of the RegARIMA model and considering the statistical significance of the model coefficients and their autocorrelations.
2. BIC: Measures the goodness of fit of the model to the data.
3. Signal Extraction: Measures the model's efficacy in signal extraction using SEATS (via canonical decomposition).
4. Revisions: The magnitude of revisions when new data is available is captured by this indicator.
5. Residual Seasonality: This indicator considers statistical tests on residual seasonality after the seasonal adjustment process is performed.

To rank the models effectively, a final score is computed by appropriately combining the five indicators. Importantly, users retain the flexibility to adjust the weights assigned to each area according to their specific requirements. For instance, users could prioritize models with minimal revisions based on their preferences. Moreover, an alternative approach based on the Pareto boundary is also explored. Finally, TEAM presents the user with a selection of the best models based on the final score, enabling them to choose the most suitable model according to their needs.

Keywords

time series, seasonal adjustment, JDemetra+

Authors and Affiliations

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

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

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

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

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

INE-Spain

1. Introduction

The TEAM (Time Series Exhaustive Automatic Modelling) software is being developed in Statistics Spain to try and overcome some difficulties that arise when using the TRAMO-SEATS methodology, namely, the fact that, for some series, the users consider that, for different reasons, the (only) model provided by the automatic model identification of TRAMO-SEATS is not adequate. This issue becomes particularly acute at the annual phase of the revision policy (see (Eurostat, 2015) and (INE Spain, 2024)), when the RegARIMA model used during the last year for each series must be evaluated and changed if necessary. In some departments, such as Quarterly National Accounts, there are a lot of series to evaluate and very little time and resources to do it. If the model provided by TRAMO-SEATS is unacceptable, the manual search for a good model can be a time-consuming task. Moreover, these difficulties have been exacerbated by the COVID-19 crisis, making it more difficult to find good models for some series.

The TEAM software is an R package. TEAM uses the rjdverse packages of the JDemetra+ ecosystem, that call the core functions in JDemetra+ (JDemetra+ Reference Manual).

The idea behind TEAM is to fit many more models than the TRAMO-SEATS program or its JDemetra+ implementation do, and to rank them according to some criteria. Then a list of the best models is provided to the user to make it more likely that the user will find that at least one of those models is adequate.

The programs that implement the TRAMO-SEATS methodology consider also different low orders of the ARIMA model, but are not exhaustive, making some simplifications for the sake of speed of computation.

All these enhancements come at the cost of increasing the computing time needed to fit each time series. For this reason and thanks to the fact that the specifications calculation process is highly parallelizable, the TEAM software has been designed to take advantage of the ability of the R language to run in parallel on the current machine and even on remote machines (by creating a cluster of workers via SSH connection), depending on the capabilities of the available computational systems. The user can also tell TEAM whether to use these possibilities or run the process sequentially.

In the remainder of the paper, we will use the term specification to refer to the set that includes the fitted model and the seasonal adjustment achieved with that model.

2. Overview of the software

In this section we present a high-level description of the software.

The execution of TEAM consists of two independent steps. In the first step, the user establishes the specifications TEAM is going to try, while in the second step each of these specifications is fitted using JDemetra+, some quality indicators are computed and finally all the specifications are ranked according to their quality. Moreover, the quality of each specification is measured through different dimensions (goodness of the RegARIMA model, quality of the canonical decomposition, behaviour of the decomposition regarding the revisions, and residual seasonality). The user can adapt the scoring system by assigning weights to the different dimensions. For example, this allows the user to give preference to models that present few revisions but are also acceptable with respect to the other dimensions of quality.

The design of TEAM is modular, facilitating easy adaptation for the use of alternative rankings for specifications. For example, it can be tailored for the X11 method of seasonal adjustment instead of the current design tailored for TRAMO-SEATS.

In the first step, the scope of the specifications to be tested can be chosen by the user. For example, the user can decide to fit models both in logs and in levels, and with different seasonal and regular differencing orders. This can be useful especially for those series in which the test used by JDemetra+ to decide whether to work in logs or in levels is not very conclusive, or for cases when the number of differences to take is quite uncertain.

The user can also tell TEAM to try different holidays calendars, different significance levels and kinds of outliers, different trading day and moving holiday treatments and different orders of the ARIMA model. These specifications are defined using R functions, with a dedicated function for each type (e.g., one for outliers, one for logs/levels, one for ARIMA orders, etc.). Subsequently, these specifications are combined using two R operators, one for the union of specifications and another for expanding two groups of specifications to generate all possible combinations. This creates a simple and highly flexible system, allowing the user to easily define all the desired specifications.

Once the specifications to try have been established, TEAM fits a model for each of the given specifications using JDemetra+. Then, the specifications which don't have a canonical decomposition are discarded, since they are unacceptable from the point of view of seasonal adjustment. Also, the specifications where some (approximate) root cancellation occurs between the regular AR and MA polynomials, or the seasonal AR and MA polynomials, are discarded. The motivation for discarding these models is that the (quasi-)cancellation of roots can cause numerical problems and is avoided in TRAMO and other time series programs. Moreover, in many cases, when root cancellation is present, there is an equivalent simpler model, and this model is also fitted in TEAM. The fact that we discard models which don't have a canonical decomposition and models with root cancellation, before computing any further quality indicator, allows us to speed up significantly the process of specification ranking.

Then, for each of the non-discarded specifications, a set of quality indicators are extracted directly from JDemetra+ and are used to compute the five main indicators for each of the five quality dimensions described above, as described in more detail in the next section. Finally, all the specifications are ranked according to these quality indicators. Presently, these are combined in a unique global score for each specification, by assigning some weights adjustable by the user to each dimension. But other scoring strategies are also being explored, see section 4. Once all the specifications have been ranked, they are presented to the user, together with their global scores and the scores in each of the quality dimensions.

3. Methodology for the ranking of specifications

The key part of the program is the comparison of the different specifications. This is not an easy task, since there are multiple criteria to consider. The solution followed has been to develop a hierarchical system of indicators which are computed for each specification, and then to rank the specifications by appropriately combining the indicators with some weights (adjustable by the user) to favour some quality dimensions.

3.1 System of indicators

The approach followed in TEAM involves a three-level hierarchical set of indicators. All the indicators in all levels are standardized to take values between 0 and 1, where the higher means the better.

At the top of the hierarchy stand five first-level quality indicators. These are:

- Goodness of the RegARIMA model obtained.
- BIC
- Quality of the canonical decomposition obtained by SEATS.
- Behaviour of the decomposition regarding the revisions.
- Residual seasonality.

Each first-level indicator is obtained from a set of second-level indicators by taking the minimum of all of them. Each second-level indicator is obtained from one or several third-level indicators using also the minimum. The third-level indicators can be directly computed from the JDemetra+ output.

The use of the minimum is an attempt to penalize specifications that perform poorly with respect to any of the indicators.

We now describe the computation of each first-level indicator in detail.

3.1.1 Goodness of the RegARIMA model

It is computed as the minimum of three second-level indicators:

- **Residual tests**: this indicator is obtained as the minimum of several third-level indicators which are computed as the p-values of statistical test on the residuals of the RegARIMA models. These tests include normality tests, independence (no autocorrelation tests, linearity tests, tests for seasonal spectral peaks, tests for trading days spectral peaks, and out-of-sample tests.
- **Significance of the RegARIMA parameters**: to compute this indicator first the p-values of the tests for statistical significance of the coefficients of the RegARIMA model are transformed, to make them comparable to those of the previous point, and then the minimum is taken.
- **Correlations of the ARIMA parameters**: this indicator considers the existence of large correlations between the ARIMA parameters. A value between 0 and 1 is assigned, where lower values mean higher correlations.

3.1.2 BIC

We use the BIC as a first-level indicator, which allows us to compare the different specifications in terms of goodness of model fit. The BIC considers not only the goodness of fit of the model, but also the number of parameters used in the model. Therefore, this criterion tends to favour models with less parameters (in our case, less regressors and lower ARIMA orders).

3.1.3 Quality of the SEATS signal extraction

It is defined as the minimum of four second-level indicators that measure the performance of SEATS:

- **Model-based tests**: this indicator compares the empirical distributions of the components produced by the canonical decomposition with the theoretical distributions of their estimators, and it is considered a measure of the quality of the estimated components based on the canonical decomposition. This indicator is defined as the minimum of the p-values of the tests regarding the variance and autocorrelation of the components and the cross-correlations between the components.

- **Canonical decomposition**: this indicator is a measure of how far a model is from a model that does not admit a canonical decomposition. It is computed as the variance of the irregular component divided by the variance of the innovations of the linearized series.
- **Significant seasonality**: this indicator measures the number of periods which present significant seasonality in the central part of the series (historical), the last year (current) and the first year ahead forecasts.
- **Final error**: this indicator considers the variance of the final error estimator, measured in units of the variance of the innovations in the linearized series.

3.1.4 Behaviour of the decomposition regarding the revisions

The minimum of three second-level indicators is used, these are:

- **Speed of convergence**: this indicator measures the speed of convergence of the concurrent estimator towards the historical estimator.
- **Revision error**: this indicator considers the revision error variance of the concurrent estimator, measured in units of the variance of the innovations in the linearized series.
- **Revision history**: this indicator considers the magnitude of the revisions for each period in the last four years.

3.1.5 Residual seasonality

This indicator evaluates the quality of the seasonally adjusted series by taking the minimum of two second-level indicators, which measure its bias and the presence of residual seasonality in it:

- **Bias in yearly totals**: the bias in the seasonally adjusted series is evaluated, computed as the maximum of the differences between the yearly means of the seasonally adjusted series and the raw series.
- **Residual seasonality tests**: this second-level indicator is computed as the minimum of the p-values obtained with several residual seasonality tests applied both to the seasonally adjusted series and to the irregular component.

3.2 Ranking of the specifications

Once we have computed the five indicators, the remaining question is how to proceed, since all of them are important and it is quite unlikely that one of the specifications will perform better than the others across the five indicators.

A two-step approach is used.

Normalization of the indicators: For each one of the five indicators, we consider $(x_{(1)}, x_{(2)}, \dots, x_{(N)})$, the vector with that indicator for all N specifications, sorted in ascending order. The normalized indicator for specification $i \in \{1, \dots, N\}$ is defined as:

$$\hat{x}_i := \frac{\sum_{j=1}^{i-1} \left(x_{(j+1)} - x_{(j)}\right)^2}{\sum_{j=1}^{N-1} \left(x_{(j+1)} - x_{(j)}\right)^2}$$

This definition provides a normalized indicator equal to 0 for the specification with the lowest value in the indicator and equal to 1 for the specification with the highest value in the indicator. Moreover, the value of the normalized indicator for a specification depends not only on its relative position in relation to the other specifications, but also on its distance relative to the other specifications.

Computation of the final score: Including a new index $j \in \{1, \dots, 5\}$ for the number of indicator, we now have the normalized indicators $\hat{x}_{i,j}$ for each specification i and indicator j.

The final score S_i for specification i is computed as:

$$S_i = \sum_{j=1}^{5} w_j \log\left(\hat{x}_{i,j}\right)$$

where the w_j is the weight assigned to each indicator, that can be chosen by the user. Typically, the weights will be positive real numbers. The bigger the final score, the better the specification is considered.

In the last formula, the logarithm is used to penalize a very low value in any of the indicators, avoiding so heavily unbalanced specifications, like one that ranks, for example, very well in one indicator, but badly in other one. As an extreme case, if a specification has one or more null normalized indicators, its final score will be $-\infty$.

4. Future work

The methodology explained in this paper has been thoroughly tested using the short-term statistics from Statistics Spain (Gómez et al., 2024). These tests show that, although the methodology works well in most cases, there is still room for improvement. Therefore, the next steps in the development of TEAM will be to adjust the present methodology to account for those cases where the procedure does not perform in an optimal way.

Another line for improvement is to offer, in addition to the linear ranking of the specifications explained in section 3.2, an alternative multi-objective approach in the ranking of specifications. With this perspective, we compute the Pareto front (the specifications which are not dominated by any other in at least some of the five first-order indicators) and allow the user to select one of these specifications according to their needs.

Acknowledgment

The authors want to thank the Short-term Statistics and National Accounts departments of Statistics Spain for their thoroughly testing of the methodology presented in this paper and for their valuable suggestions for its improvement.

References

C. Gómez, E. Rosa-Perez, C. Sáez Calvo, L. Sanguiao Sande, Aparicio Pérez, F., M. T. Vázquez Gutiérrez and J. F. Arranz Arauzo. Improving quality in seasonal adjustment in Short-Term Statistics using JDemetra+ regressors and TEAM R-package. 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

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/