(In)Stability of Reg-ARIMA Models for Seasonal Adjustment

Abstract

Regression models with ARIMA errors (Reg-ARIMA) are nowadays commonly used in seasonal adjustment to remove the main deterministic effects (outliers, ruptures, calendar effects) from the raw data before decomposing the corrected series into trend-cycle, seasonality and irregular.

The main seasonal adjustment programs (X-13ARIMA-SEATS, TRAMO-SEATS, JDEMETRA+) implement these models in an automatic and very user-friendly way. This facility hides in fact a real complexity and in certain cases a lack of robustness which can escape the user.

In this presentation, we draw attention to the real difficulties of implementing these models through concrete cases: the estimation of a leap year effect, the estimation of breaks and even the estimation of an ARIMA model.

Contents

List of Tables

List of Figures

Introduction

The most popular seasonal adjustment methods are undoubtedly TRAMO-SEATS (Gómez and Maravall (1996) and Maravall and Caporello (2004)), a parametric method based on ARIMA models and X-13ARIMA-SEATS (Findley and al. (1998) and Ladiray and Quenneville (2001)), a non-parametric method based on moving averages. Both methods are recommended by Eurostat and the European Central Bank (ECB) for the seasonal adjustment of economic indicators.

These methods proceed in two steps summarized in figure 1. First, they adjust the raw series from some deterministic effects (outliers, breaks, calendar effects) and, in a second step, they decompose the series thus "cleaned" into trend-cycle, seasonal component and irregular component.
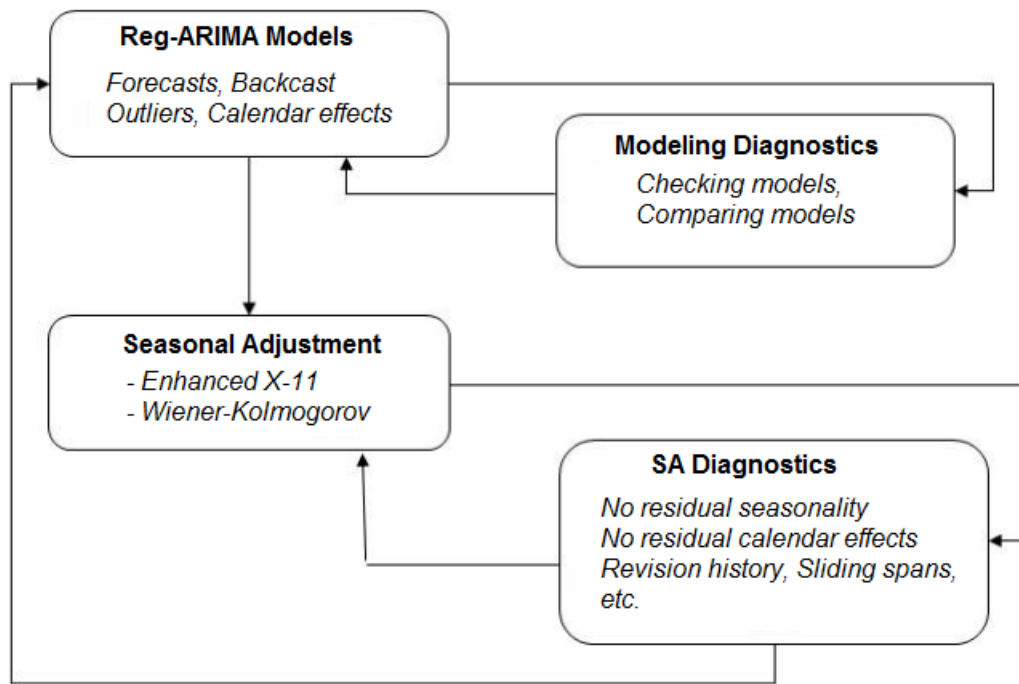
Figure 1: The seasonal adjustment process of X-13ARIMA-SEATS and TRAMO-SEATS: correction and decomposition.

The pre-decomposition stage, called pre-adjustment, is based on a "regression model with ARIMA errors" (Reg-ARIMA model). An automatic modeling algorithm (AMI) is available to the user that allows to:

- determine the mode of the model (additive or multiplicative);

- detect and correct breaks in the series;

- detect and correct potel calendar effects (trading day effects and moving holidays such as Easter);

- adjust the "residuals" of the regression with an ARIMA model;

- estimate any missing values;

- and finally to forecast and backcast the series studied.

This automatic algorithm is implemented in a very user-friendly in softwares and it always comes with a large battery of statistical tests to judge the quality of the adjustment made. A very detailed description can be found in Gómez and Maravall (1998).

Maravall and al. (2015), Maravall and al. (2016), Maravall (2018) assess the performances of this AMI. Their results are laudatory for this tool but they consider, on the one hand, a modeling at a given moment and they validate, on the other hand, the method by the tests which finally served to build it. Nevertheless, the AMI is a first-rate **exploratory** tool, unfortunately too often used in practice as a black box.

The goal of this article is, based on simple simulations on real data and focusing on the stability of estimates, to show that these Reg-ARIMA models are complex, difficult to use, and must be handled with caution.

In the first part, we briefly remind the definition of Reg-ARIMA models and the main deterministic components used in this work. The second part of the study is devoted to the estimate of the leap year effect in the industrial production index (IPI) and the turnover index (ITI) series of countries of the European Union. The third part focuses in the estimation of four kinds of outliers — additive outliers, transitory change, level shift and seasonal outliers — in the IPI series. The fourth part looks at the estimates of the ARIMA model itself.

This article does not **prove** anything: it is settle for **showing** on concrete examples, that the modelization Reg-ARIMA suffers, like number of parametric methods, of a certain lack of robustness. The use of these models, also very powerful, requires both a good knowledge of the underlying theory and its limits, and great caution as well summarized Findley and McElroy (2018):

> "*ARIMA Model-Based Seasonal Adjustment (AMBSA) should not be a black box procedure to its users because default software procedures are sometimes seriously inadequate. Also user decisions regarding software options and model choice can strongly impact the results obtained, for better or worse. A seasonal adjuster who understands the basic facets of the method and some of its diagnostics, as outlined and then detailed in this document, will have a greater capacity to obtain successful adjustments*".

The proper use of these models also requires a bit of common sense. Thus, and by way of example, it is certainly absurd to try to correctly estimate a leap year effect with a model on a series of less than 9 years old. To estimate this effect, you will have observed only two leap years, which is very little.

1 Reg-ARIMA models

1.1 Definition

Let z_t an observed time series, monthly ($s = 12$) or quaterly ($s = 4$) and if necessary log-transformed. X-13Arima-Seats and Tramo-Seats adjust z_t a model:

$$
\begin{aligned}
z_t &= y_t'\beta + x_t \\
\phi(B)x_t &= \Theta(B)a_t,
\end{aligned}
\tag{1}
$$

Where:

- $y_t'\beta$ is the part of the regression that takes into account the deterministic components of the series;

- x_t, the "residuals", follows an ARIMA model $(p,d,q)(P,D,Q)_s$;

- $\phi(B)$ and $\Theta(B)$ are polynomials of the lag operator B $(B^j z_t = z_{t-j})$;

- and a_t is a white noise.

For seasonal series, a multiplicative specification is used:

$$\varphi_r(B)\varphi_s(B^s)\nabla^D\nabla_s^{D_s}x_t = \theta_r(B)\theta_s(B^s)a_t, \tag{2}$$

where:

- $(\nabla^D, \nabla_s^{D_s})$ is the differentiation that leads to stationarity, with $\nabla^D = (1-B)^D$ and $\nabla_s^{D_s} = (1-B^s)^{D_s}$,

- $\varphi_r(B)$ and $\varphi_s(B^s)$ are the autoregressive polynomials associated with the regular and seasonal stationary components, order P_r and P_s:

$$\begin{aligned}
\varphi_r(B) &= 1 + \varphi_{r,1}(B) + \cdots + \varphi_{r,P_r}(B^{P_r}) \\
\varphi_s(B^s) &= 1 + \varphi_{s,1}(B^s) + \cdots + \varphi_{s,P_s}(B^{sP_s}),
\end{aligned}$$

- $\theta_r(B)$ and $\theta_s(B^s)$ are the moving average polynomials associated with the regular and seasonal components, order Q_r and Q_s:

$$\begin{aligned}
\theta_r(B) &= 1 + \theta_{r,1}(B) + \cdots + \theta_{r,Q_r}(B^{Q_r}) \\
\theta_s(B^s) &= 1 + \theta_{s,1}(B^s) + \cdots + \theta_{s,Q_s}(B^{sQ_s}),
\end{aligned}$$

- and a_t is a white noise.

In the automatic algorithm implemented by X-13ARIMA-SEATS and TRAMO-SEATS, the orders of the polynomials are constrained: P_r and $Q_r \in [\![0, 3]\!]$, $D \in [\![0, 2]\!]$ and D_s, P_s and $Q_s \in [\![0, 1]\!]$, which corresponds to 384 possibles ARIMA models.

1.2 The deterministic componets of the regression

In seasonal adjustment, we usually focus on two important components: calendar effects, which must be suppressed in the series "seasonally adjusted and working days adjusted" (SA-WDA) and the presence of breaks that may hinder the estimate of the seasonal component.

1.2.1 The calendars effects

The different calendars effects, the associated models and the problems related to their detection and estimation are detailed in Ladiray (2018). Here we focus on the so-called "working days" effects linked to the fact that the composition in days of the month — the numbers of Mondays, Tuesdays, ..., Sundays — varies over time.

Following the notations of Findley and al. (1998), we assume that the day number j of the week has an effect α_j where, for example, $j = 1$ means Monday, $j = 2$ Tuesdays, ... and $j = 7$ means Sunday. For instance, each α_j represents the average sales of day j. If N_{jt} number of days j during the month t, the length of the month t is then $N_t = \sum_{j=1}^{j=7} N_{jt}$ and the cumulative effect for this month (the total sales of the month) will be: $TD_t = \sum_{j=1}^{j=7} \alpha_j N_{jt}$.

A first idea to detect and estimate the working days effects in a series is to explain the values of the series by the seven variables N_{jt}. However, these regressors are naturally seasonal (there are on average more Mondays in January than in February) and strongly correlated.

A different but equivalent specification makes it possible to resolve these problem to a large extent. The average daily effect, average saled for a day, is $\bar{\alpha} = \sum_{j=1}^{7} \alpha_j / 7$. By construction $\sum_{j=1}^{7} (\alpha_j - \bar{\alpha}) = 0$, and we have:

$$
\begin{aligned}
\sum_{j=1}^{7} \alpha_j N_{jt} &= \bar{\alpha} N_t + \sum_{j=1}^{7} (\alpha_j - \bar{\alpha}) N_{jt} \\
&= \bar{\alpha} N_t + \sum_{j=1}^{6} (\alpha_j - \bar{\alpha}) (N_{jt} - N_{7t}).
\end{aligned}
$$

Thus, the cumulative effect of the month is composed of an effect directly related to the length of the month and a net effect of each day of the week. Since the quantity $\bar{\alpha} N_t$ is seasonal by nature (January always has more days than February), we use the equality: $\bar{\alpha} N_t = \bar{\alpha} N_t^* + \bar{\alpha} (N_t - N_t^*)$, where N_t^* is the average of the month of length t. In other words, N_t^* is equal to 30 or 31 if the month in question is not a February and 28.25^1 otherwise. The second term of equality is therefore null except for the month of February. The use of the contrasts variables thus allows seasonal adjustment of the regressors working days.

Current versions of X-13ARIMA-SEATS and TRAMO-SEATS use the following Reg-ARIMA model to estimate trading days effects:

$$
z_t = \beta_0 LY_t + \sum_{j=1}^{j=6} \beta_j (N_{jt} - N_{7t}) + x_t, \tag{3}
$$

where LY_t is the leap year (LY) regressor equals to:

$$
LY_t = \begin{cases} 0.75 & \text{if } t \text{ is a February during leap years} \\ -0.25 & \text{if } t \text{ is a February during non leap years} \\ 0 & \text{otherwise} \end{cases}
$$

1.2.2 Outliers and breaks

The different outliers, the associated models and the problems related to their detection and estimate are described in detail in Mehrhoff (2018).

In the continuation, we will will focus on the following breaks:

- An additive outlier (AO), a break affecting only one observation t_0, one month. AO are modeled with the following regressor:

$$
AO_t = \begin{cases} 1 & \text{if } t = t_0 \\ 0 & \text{if } t \neq t_0 \end{cases}
$$

- A level shift (LS), representing a permanent break in the level of the series from the time t_0. LS are modeled with the following regressor:

$$
LS_t = \begin{cases} -1 & \text{if } t < t_0 \\ 0 & \text{if } t \geq t_0 \end{cases}
$$

- A transitory change (TC), a break on a given time t_0 that decays exponentially over the following periods. TC are modeled with the following regressor:

$$
TC_t = \begin{cases} 0 & \text{if } t < t_0 \\ \alpha^{t-t_0} & \text{if } t \geq t_0 \end{cases}
$$

where α, which is the speed back to normal, is by default equals to 0.7.

---

[1] In fact, the true value would be 28.2425 as shown in the appendix (A).

- A seasonal outlier (SO), an abrupt change in the seasonal pattern of a given month. SO are modeled with the following regressor:

$$SO_t = \begin{cases} 0 & \text{if } t \geq t_0 \\ 1 & \text{if } t < t_0 \text{ and } t \text{ is the same month/quarter of } t_0 \\ -1/(s-1) & \text{otherwise} \end{cases}$$

Figure 2 gives a graphical representation of the regressors associated with these four breaks.

AO and LS are relatively common in economic series. SO and TC are much rarer and often difficult to explain from an economic point of view. Other types of breaks are sometimes considered: ramps (RP), temporary level changes TLS...
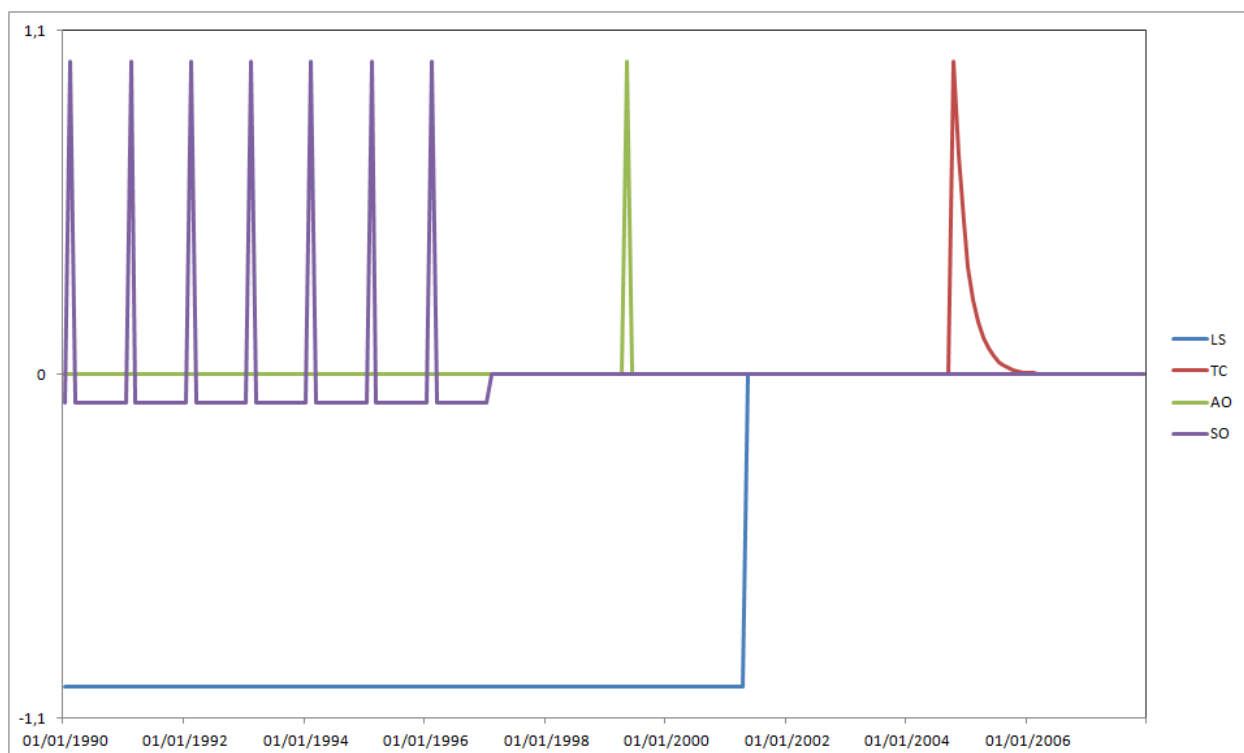

Figure 2: Regressors associated with some kind of breaks (AO, LS, TC, SO).

1.3 An example X-13Arima-Seats

As an example, figure 3 shows the result of an automatic modelling on the IPI of the French manufacturing industry, from January 2006 to April 2017. Computations were made with the software JDemetra+[2] version 2.2.1, using the Reg-ARIMA module implemented in X-13Arima-Seats and and the default pre-defined specification "RSA5c", see Grudkowska (2017) for more information.

All the statistical indicators associated to the computations are "green" (i.e.: considered "good" by JDemetra+) and the final model shows:

---

[2]JDemetra+ is the seasonal adjustment software officially recommended to the members of the European Statistical System (ESS) and the European System of Central Banks. In particular, it implements the methods X-13Arima-Seats and Tramo-Seats.

Table 1: Comparison of the results of the AMI for X13-Arima-Seats and Tramo-Seats on the 205 IPI series (6 trading days regressors and leap year regressors).

Result	Model	Trading day diagnostic?
Different	84	22
Identical	121	183
Total	205	205

- A strong working days effect with a negative estimate coefficient expected for the weekend;

- A leap year effect;

- A more surprising Easter effect;

- An additive outlier in May 2011 when the production seems to have been much stronger than expected.

When these deterministic effects are removed, the series follows an Airline model $(0, 1, 1)(0, 1, 1)$.

This simple example highlights some of the limitations of a "blind" use of the AMI: Why would industrial production fall in the week before Easter? The use of the "RSA5c" option is equivalent to specify a specific Reg-ARIMA model and in particular to consider that the IPI is affected by the Easter holiday. If this hypothesis could be eventually accepted for some sectors (lamb meat, chocolate production), it seems harebrained at the global level. On the other hand, some sectors of the retail trade could have an Easter effect (sale of chocolate, lamb, flowers).

The AMI, which automatically implements the Box and Jenkins method, helps the analyst to formulate, **to specify**, a model by revealing to him the possible presence of outliers and breaks, of calendar effects and by proposing an ARIMA model for the residuals of this model. This algorithm is particularly useful for starting an analysis, especially when one has to deal with tens or hundreds of series. It is also sometimes used in an annual seasonal adjustment campaign to check if the models used in production could not be substituted by others.

It must be understood and kept in mind that this tool is exploratory. Despite its exceptional effectiveness, it is not a confirmatory tool and its results have no character of absolute truth. It's easy to show that:

- In practice, it often happens that the model proposed by the AMI is modified to: include atypical points or breaks, change the system of regressors for working days, even modify the ARIMA model, etc.

- It is not uncommon that Seats reject the ARIMA model proposed by Tramo to choose a decomposable model.

- The AMI algorithms of X13-Arima-Seats and Tramo-Seats are different and give different results. By way of illustration, both algorithms were appied on 205 IPI's series: the ARIMA model is not the same in 60% of the cases and the diagnosis on the significance of the working days set of regressors is different for 10% of the series (see table 1).

- These algorithms use many statistical tests and are very sensitive to preprogrammed threshold values. Thus, changing the default value of the significance test for atypical points can have very important consequences for the final proposed model.

We will now illustrate some problems related to a "blind use" of this algorithm by focusing on the stability and relevance of the Reg-ARIMA model estimates.
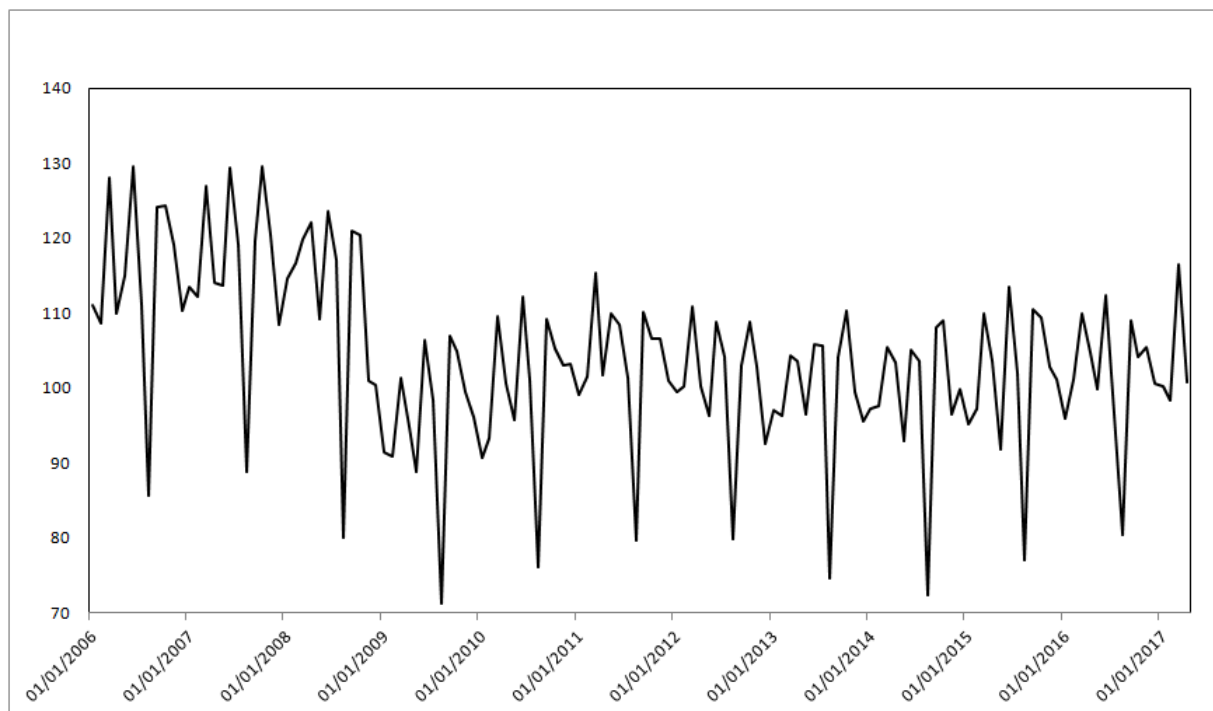
Summary

Estimation span: [1-2006 - 4-2017]
136 observations
Series has been log-transformed
Trading days effects (7 variables)
Easter [8] detected
1 detected outlier

Regression model
Trading days

	Coefficients	T-Stat	P[|T| > t]
Monday	0,0116	3,42	0,0009
Tuesday	0,0126	3,66	0,0004
Wednesday	0,0068	1,98	0,0501
Thursday	-0,0014	-0,41	0,6852
Friday	0,0139	4,07	0,0001
Saturday	-0,0197	-5,63	0,0000
Sunday (derived)	-0,0238	-6,66	0,0000

Joint F-Test = 53,00 (0,0000)

Arima model
[(0,1,1)(0,1,1)]

	Coefficients	T-Stat	P[|T| > t]
Theta(1)	-0,1934	-2,09	0,0393
BTheta(1)	-0,7322	-9,77	0,0000

Leap year

	Coefficients	T-Stat	P[|T| > t]
Leap year	0,0179	1,64	0,1037

Easter [8]

	Coefficients	T-Stat	P[|T| > t]
Easter [8]	-0,0210	-3,04	0,0029

Outliers

	Coefficients	T-Stat	P[|T| > t]
AO (5-2011)	0,1284	7,30	0,0000

Figure 3: Automatic modeling of the IPI of the French manufacturing industry.

2 Estimate of a leap year effect

As shown by the equation 3), the leap year effect is a component of the working day effect generated by the composition in days of the month, variable from one year and one month to the other.

The presence of trading days regressors in the Reg-ARIMA model implies, on the one hand, that these effects exist, and ,on the other hand, that they are largely independent of the ARIMA model[3], and finally that they are constant over time.

Apart from the assumption of independence, these hypothesis can only be validated putting the series back to their real context. For that matter, the seasonal adjustment guidelines, Eurostat (2015), specifies that:

"The calendar adjustment should be done for those time series for which there is an economic rationale for the existence of calendar effects and statistical evidence."

In this study we assume that industrial production and turnover indices are impacted by calendar effects. This is undoubtedly true because these effects result from various laws or branch agreements: compulsory weekly breaks, usually on Sundays, legal holidays, etc. But on the one hand, there are many possible adjustments and, on the other hand, industries adapt their production process to make it less sensitive to holidays, weekends, holidays, etc. In practice, it is also easy to check that the weekend is a period of low industrial production and that Saturday is a very favorable day for turnover of retail businesses.

Likewise, it is obvious that there is a leap year effect in all series that measure a monthly activity as a sum of daily activities[4]. One extra day can only increase the production or the turnover, and whatever the model Reg-ARIMA and associated tests say: on average, this increase will be of $(29/28 - 1) = 0.0357$ so around 3.6%. Bell (1992) explains this figure for a multiplicative model and even proposes in this case to make a prior adjustment for the leap year effect rather than to estimate it in a Reg-ARIMA model, which gave rise to the *autoadjust* option of X13-ARIMA-SEATS.

2.1 Methodology

In this study on the quality of the estimates of the leap year effect, we use the monthly industrial production and turnover indices in volume, at the 2, 3 or 4-digit levels of the NACE rev 2. To study the convergence of the estimates, we only keep series of more than 12 years. In total, 2198 series are used in the simulation[5].

For each series, we proceed as follows:

1. In a first step, the decomposition model, the atypical points, the breaks, the working days effects and the ARIMA model are determined on the whole series. The Reg-ARIMA model is therefore determined.

2. In a second step, the Reg-ARIMA model is re-estimated on the first 48 observations of the series, to be certain to observe at least one leap year.

---

[3]Is to ensure as much as possible this independence that the regressors are seasonally adjusted (this is one of the advantages of considering contrasts).

[4]However, it is possible that this economic "evidence" is disturbed by the statistical collection itself. Thus, in the context of the IPI, some productions are, or have been, measured in whole production units. In the case of the manufacture of aircraft or steamers, it is certain that the leap year effect would not be discernible because of a bias related to the truncation of the measure at monthly level. This kind of problem is less common since worked or billed hours are increasingly used.

[5]These series are available on the Eurostat website, in the tables *sts_intv_m*, *sts_setu_m*, *sts_sepr_m* and *sts_inpr_m*. http://ec.europa.eu/eurostat/data/database.

3. Finally, the process is repeated by adding an observation each time, which gives new estimates of the parameters. For a 13-year series, we will therefore have $12 \times 13 - 48 = 108$ estimates of the coefficient of the leap year variable.

These simulations make it possible to study the convergence of the leap year coefficient. In this study, we will say that the estimate of the leap year coefficient has converged when it remains positive, statistically significant and not significantly different from the last estimate (obtained estimating the model with all the data).

Other specifications are of course possible for this simulation: shortening the initial estimation period, not fixing the model, etc. We have conducted different experiments that lead to similar results. Only the results of this simulution with a fixed Reg-ARIMA model are presented here.

2.2 Results

In the estimation of this leap year effect, a Reg-ARIMA model is naturally confronted with several difficulties:

- The lack of data: to identify the effect of a leap year, we must observe leap years... and enough.

- The difficulty of isolating each component, we remind that they are not observed directly, especially when the series is very noisy.

The main results are well summarized in the figures 4 and 5; in these figures, the shaded areas represent the 90% and 95% confidence intervals.

- In both cases, we see that the leap year coefficient takes time to converge and the values of the first years are particularly erratic.

- The first leap year is observed here in 1992, and the others in 1996, 2000, 2004, etc. It is only from 2004 that the coefficient converges for the series RF0610; convergence is much slower for the series RF1391.

- The estimated values for the RF1391 series are not economically justifiable: how could one more day in a month increase production, all other things being equal, from 7% to 14%, or even 19%?

The convergence indicators have been calculated *in fine* on 624 series. The figures 6 and 7 summarize the results of the simulation, in the form of "violins": the violins are variants of the boxplots. The shape of the box represent the distribution and the horizontal lines are the corresponding quartiles.

In terms of time needed to obtain a stable estimate, the figure 6) suggests that:

- For three-quarters of the series, it takes at least 10 years for the leap year effect to converge. It is a simple question of common sense: below 3 observations, it is most often impossible to measure with fiability a phenomenon, here the difference between the months of February of normal years and leap years.

- The nature of the model, additive or multiplicative, has little influence on the time convergence.

- Convergence seems faster for turnover series than for industrial production series.

Regarding the value of the coefficient, the figure 7 suggests that:
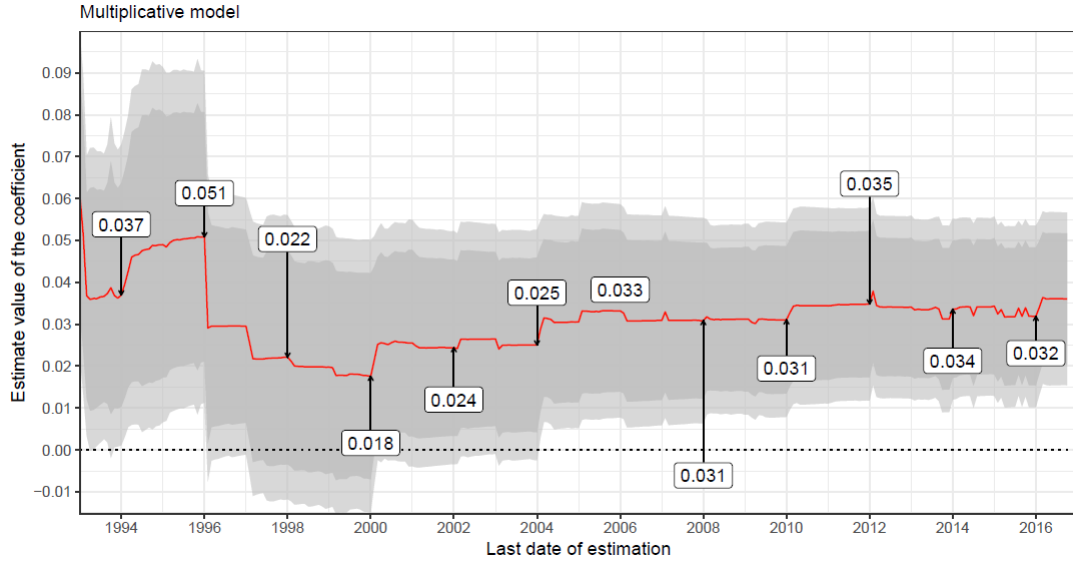
Figure 4: Successive estimates of the leap year coefficient for the series RF0610 (extraction of crude petroleum).
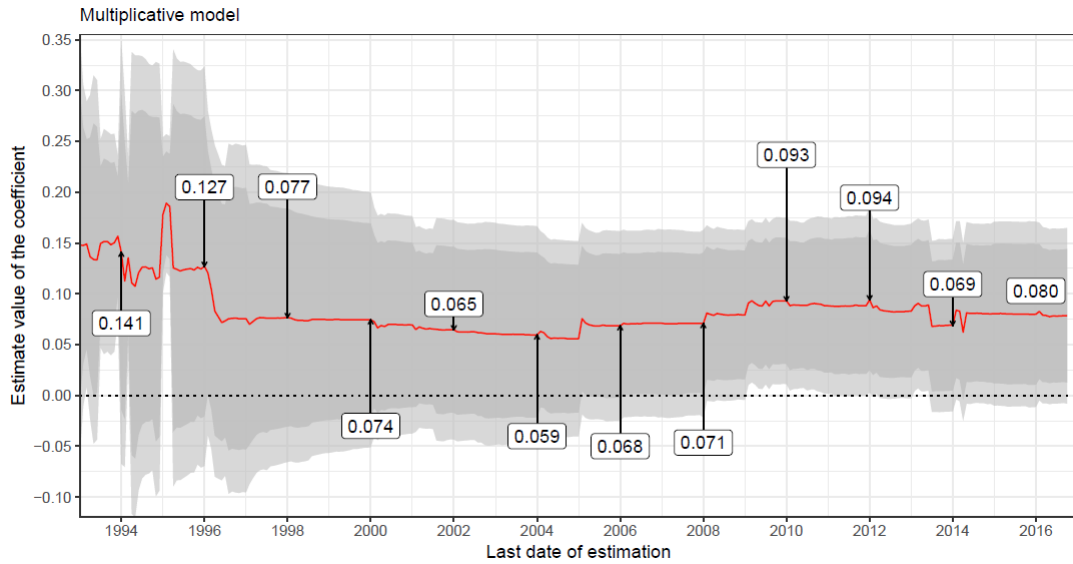

Figure 5: Successive estimates of the leap year coefficient for the series RF1391 (manufacture of knitted and crocheted fabrics).

- For multiplicative models, turnover indices (ITI) converge to a median value close to the expected value (3.5%) as Bell (1992) underlines. However, for the IPI, the median value is higher, close to 0.45 point.

- For the additive models, the ITI series converge in median to 3 index points, which approximately corresponds to the expected value since the indices are not in general very different from 100. On the other hand, for the IPI, this median value is higher, close 4.5 points.

- For many series, about 20%, the estimate value seems excessive: beyond 6 points and sometimes much more.
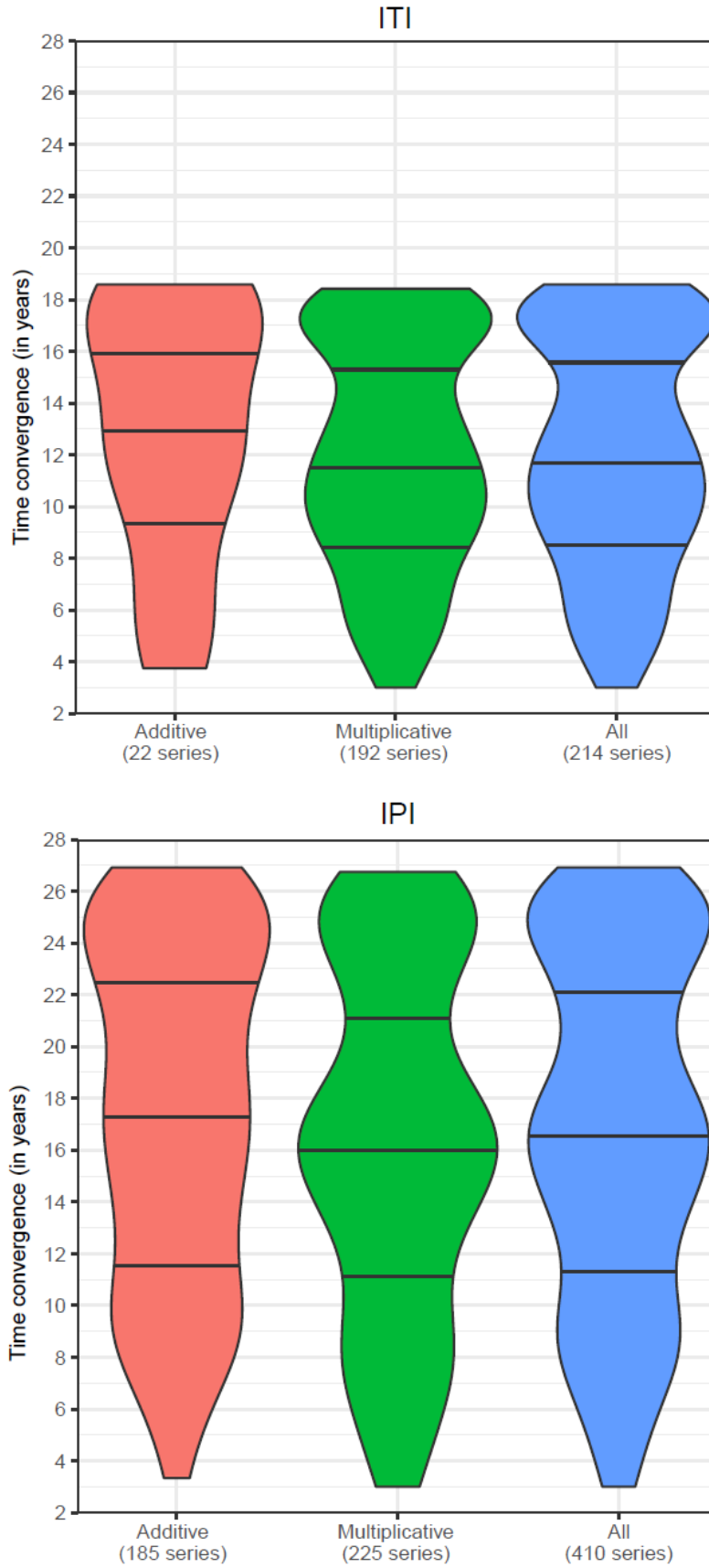
Figure 6: Time convergence (in years) of the leap year coefficient (study on 624 series).

Reading note: on the ordinate, the probability density of the time convergence of the estimate of the leap year regressor. The horizontal lines represent the associated quartiles (minimum, first quartile, median, third quartile and maximum).
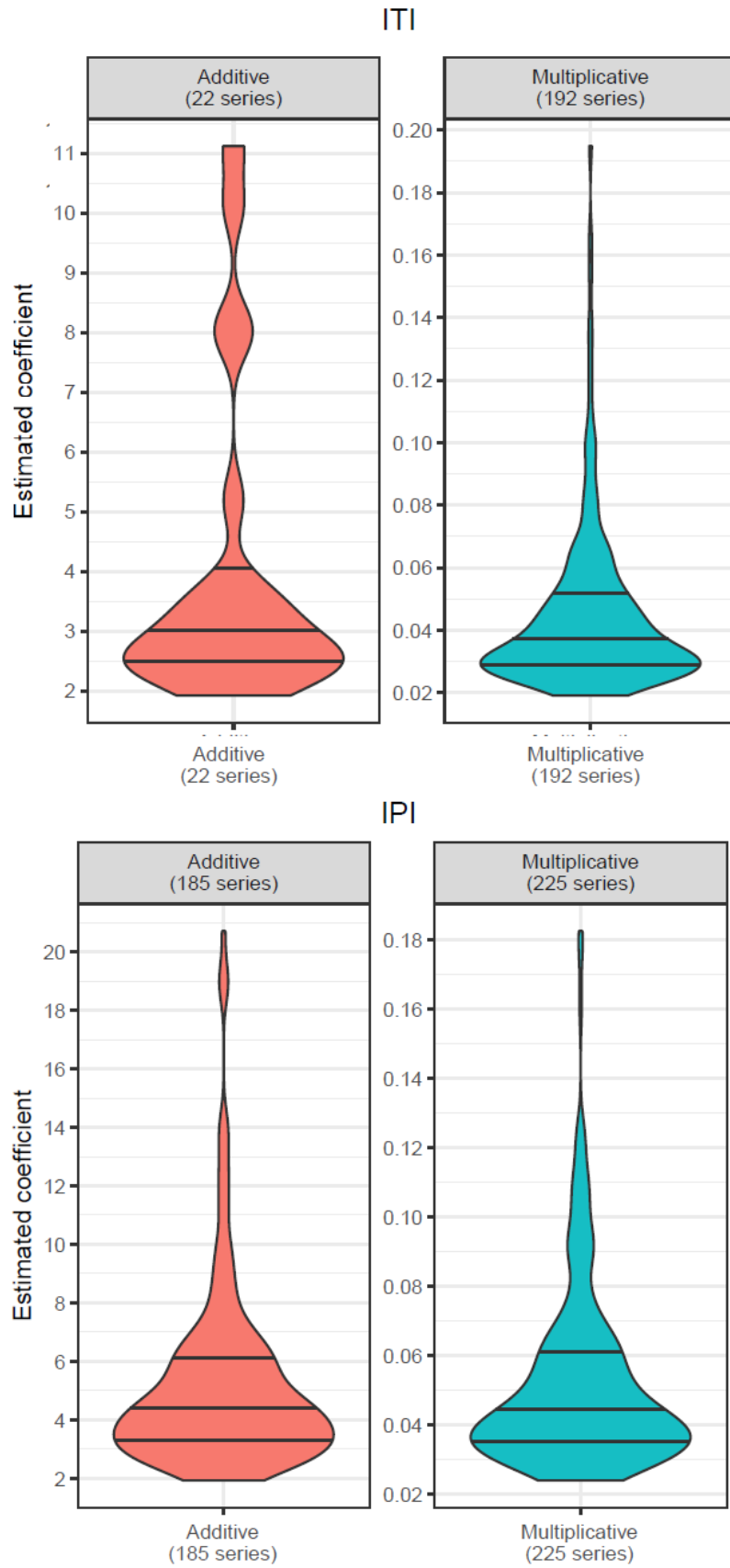
Figure 7: Final estimate of the leap year effect (study on 624 series).

Reading note: on the ordinate, the probability density of the estimate of the leap year regressor. The horizontal lines represent the associated quartiles (minimum, first quartile, median, third quartile and maximum).

In conclusion of this first experiment, it seems illusory to want to seriously estimate a "leap year effect" with series that contain less than 4, that is to say in the best case of less than 13 years.

In addition, the question arises whether it is better to correct *a priori* this effect by estimating it equal to 3.5%. To provide some answers to this question, we analyzed all 2198 series in two ways:

1. Model 1: correcting before the Reg-ARIMA modeling the raw series by a corrective coefficient equal to: $\alpha_t = \begin{cases} \frac{28.25}{29} & \text{if } t \text{ is a February during leap years} \\ \frac{28.25}{28} & \text{if } t \text{ is a February during non leap years} \\ 1 \end{cases}$

2. Model 2: leaving the Reg-ARIMA model to estimate the leap year effect with the regressor: $LY_t = \begin{cases} 0.75 & \text{if } t \text{ is a February during leap years} \\ -0.25 & \text{if } t \text{ is a February during non leap years} \\ 0 & \text{otherwise} \end{cases}$

The overall quality of the adjustments is assessed with the corrected Akaike information criteria ($AICC$): the model 1 is then considered better than the model 2 if ($AICC1 \leq AICC2$). The adjustments are made, as before, by adding one observation at a time. The figure 7 shows the change of the percentage of series for which ($AICC1 \leq AICC2$), as a function of time and the decomposition model.

The results of this experiment are unequivocal:

- For additive models, manual correction is better in about 85% of the cases.

- For multiplicative models, manual correction is better in about 80% of the cases when the series is long enough. The percentage drops to 70-75% for shorter series.
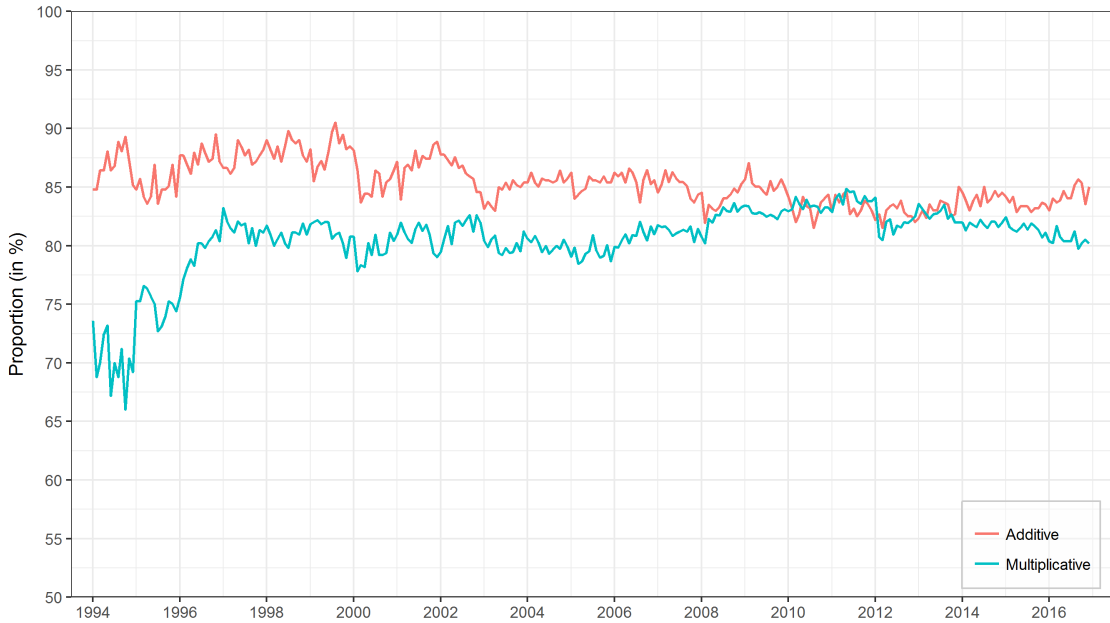

Figure 8:   Change of the percentage of series for which the manual leap year preadjustment is better (in terms of $AICC$) than the Reg-ARIMA pre-adjustment (study on 2198 series).

3 Breaks estimate

In this section, we study the estimate of outliers and breaks defined in the subsection 1.2.2, and more precisely additive outliers (AO), level shifts (LS), transitory changes (TC), seasonal outliers (SO).

3.1 Methodology

We use the monthly volume indices of the industrial production of the countries of the European community at the four-digit level of NACE rev 2. We only keep the first 12 years of the series to study the convergence of the estimates. The methodology chosen is very similar to that followed in the previous section. However, to compare the results more easily, the series are corrected from the atypical points in the year of the introduction of the break and they are all rebased to 100 the month of its introduction.

1. The decomposition model, atypical points, breaks, trading day effects and the ARIMA model are identified and estimated for the entire series (12 years). The Reg-ARIMA model is therefore determined.

2. The raw series is corrected for atypical points in the year of introduction of the break and is rebased to 100 at the month of the introduction of the break (49th observation). These operations do not modify the Reg-ARIMA model, even if, in the case of an additive schema, the rebasing implies a modification of the estimated coefficients which are multiplied by the rebasing coefficient.

3. We simulate a break at the 49th observation of the series, adding to the series the corresponding regressor multiplied by 10 if the model is additive or by 1.1 if it is multiplicative.

4. The regressor related to the studied break is added to the Reg-ARIMA model and its coefficient is estimated on the first 49 observations of the series by freezing all the other parameters estimated in the first step.

5. Finally, the process is repeated by adding an observation each time, which gives new estimates of the parameters. For a 12-year series, we will have $12 \times 12 - 48 = 96$ estimates of the coefficient of the regressor modeling the break.

These simulations then make it possible to study the convergence of the estimate of the coefficient of the break. In this study, we will say that the estimate of the coefficient of the break has converged when it remains less than 5% in absolute value of the last estimate.

3.2 Results

The results are shown in figures 9 and 10; in these figures, the red curve represents the estimates, the blue curve is a smoothed version, the shaded areas represent the 95% and 99% confidence intervals and the dotted line is the real value of the break.

- In both cases, the coefficient of the AO takes time to converge and the estimate values of the first years can be quite erratic.

- The estimate value can be very different from the real value.

The figures 11 and 12, as well as theassociated tables 2 and3, summarize the distributions of the time convergence and the estimated value, for 461 series and the 4 types of breaks studied:

- In general, estimates take a long time to converge.

- For additive models and 75% des séries, SO, LS and TC take about 7 to converge, 5 years for AO. In median, AO and TC take 4 years to converge and LS and SO 5 years.

- Convergence seems slower for multiplicative models since for 75% of series, AO take about 6 years to converge and SO, LS and TC about 7 years. However, in median, AO, LS, TC and SO converge in 4 years.

- In median, the estimates of the different breaks converge approximately to the expected value. But in 50% of the cases, the estimated value converges to a value around $\pm$ 30% of the real value. We can also note the presence of abnormal estimates, some of which even show declines instead of the increase modeled.
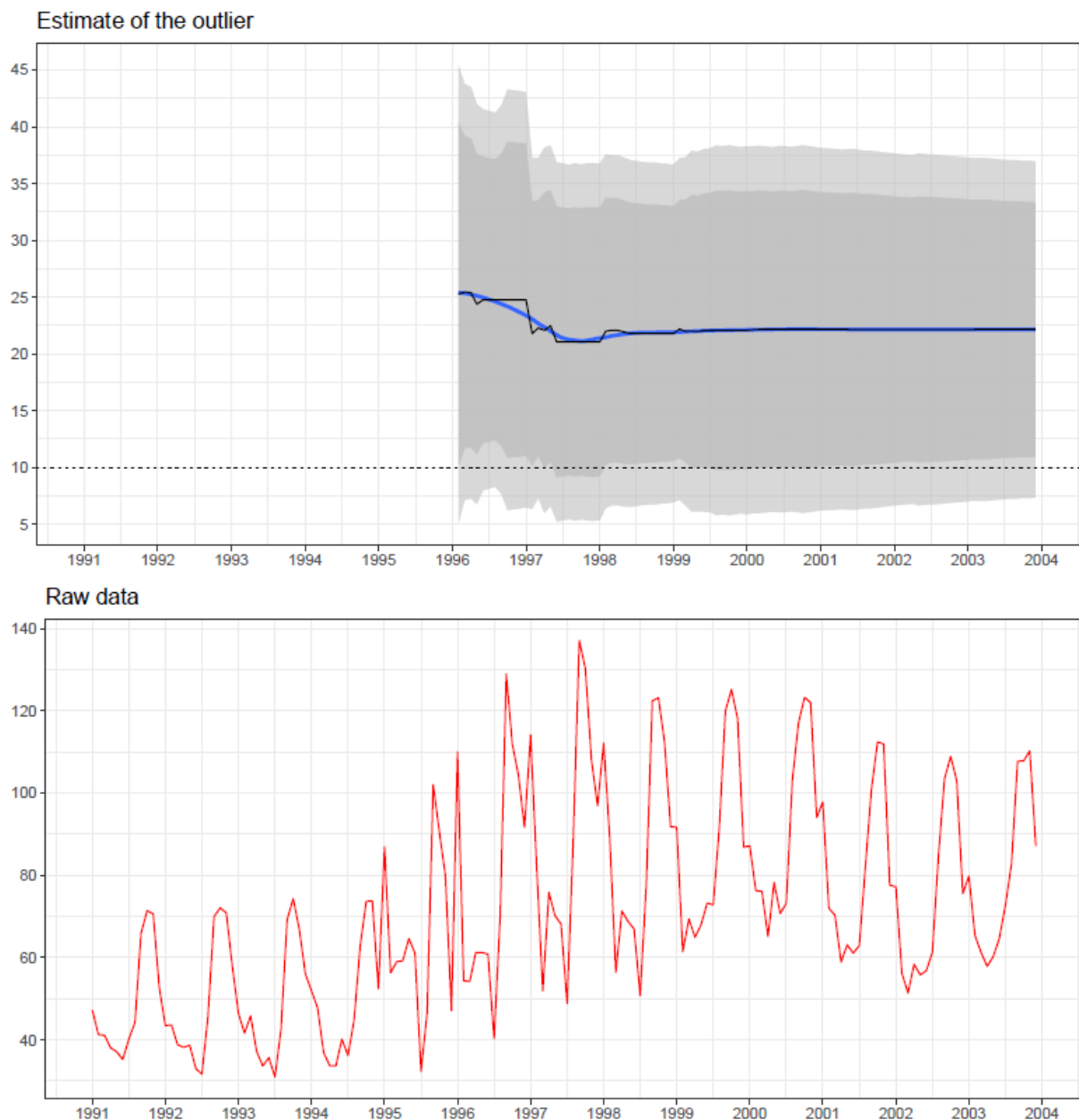
Figure 9: Raw series and change of the estimate of the additive outlier AO1996.jan for the series DE-C1062 (additive model).
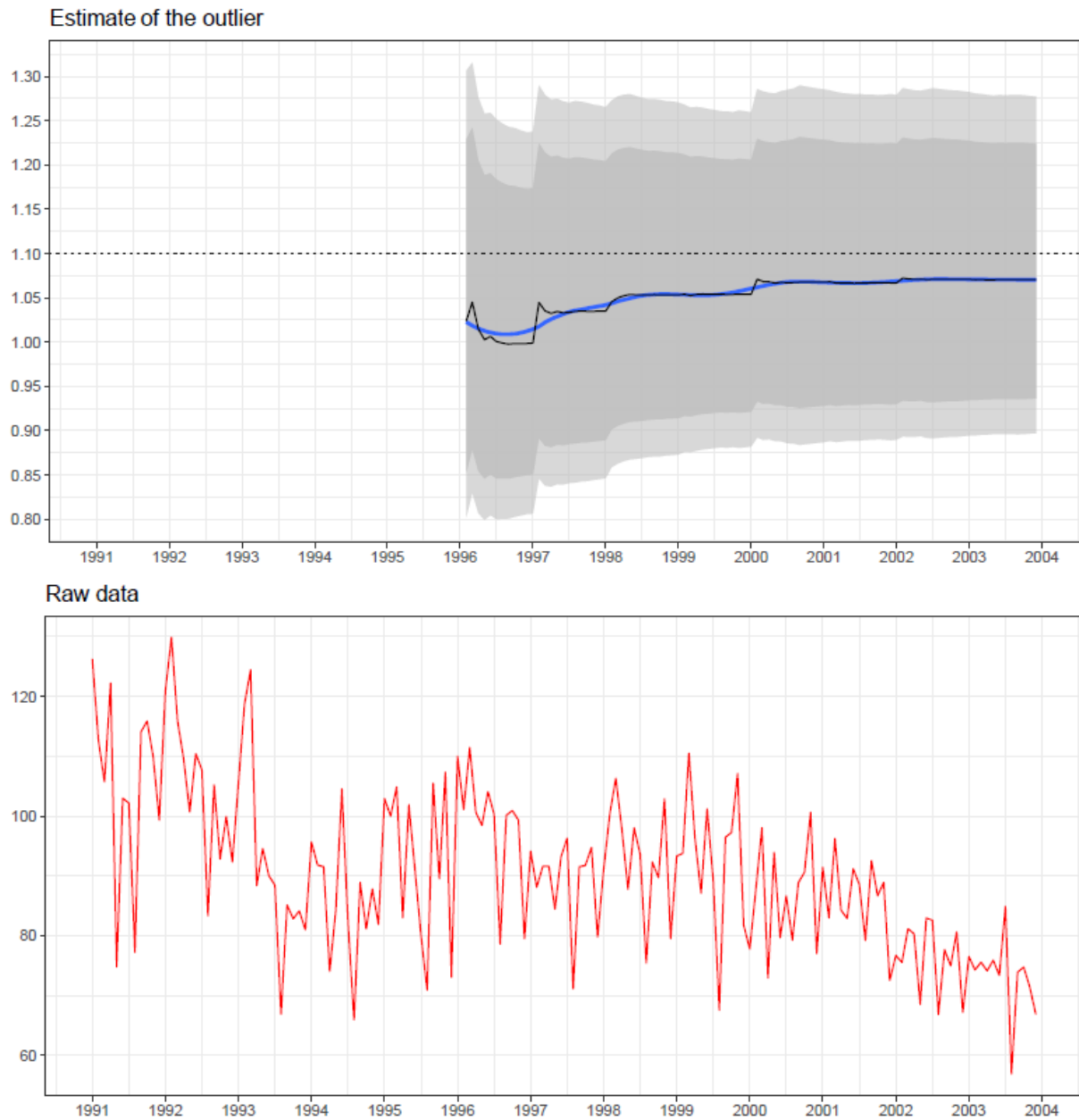
Figure 10: Raw series and change of the estimate of the additive outlier AO1996.jan for the series DE-C1086 (multiplicative model).
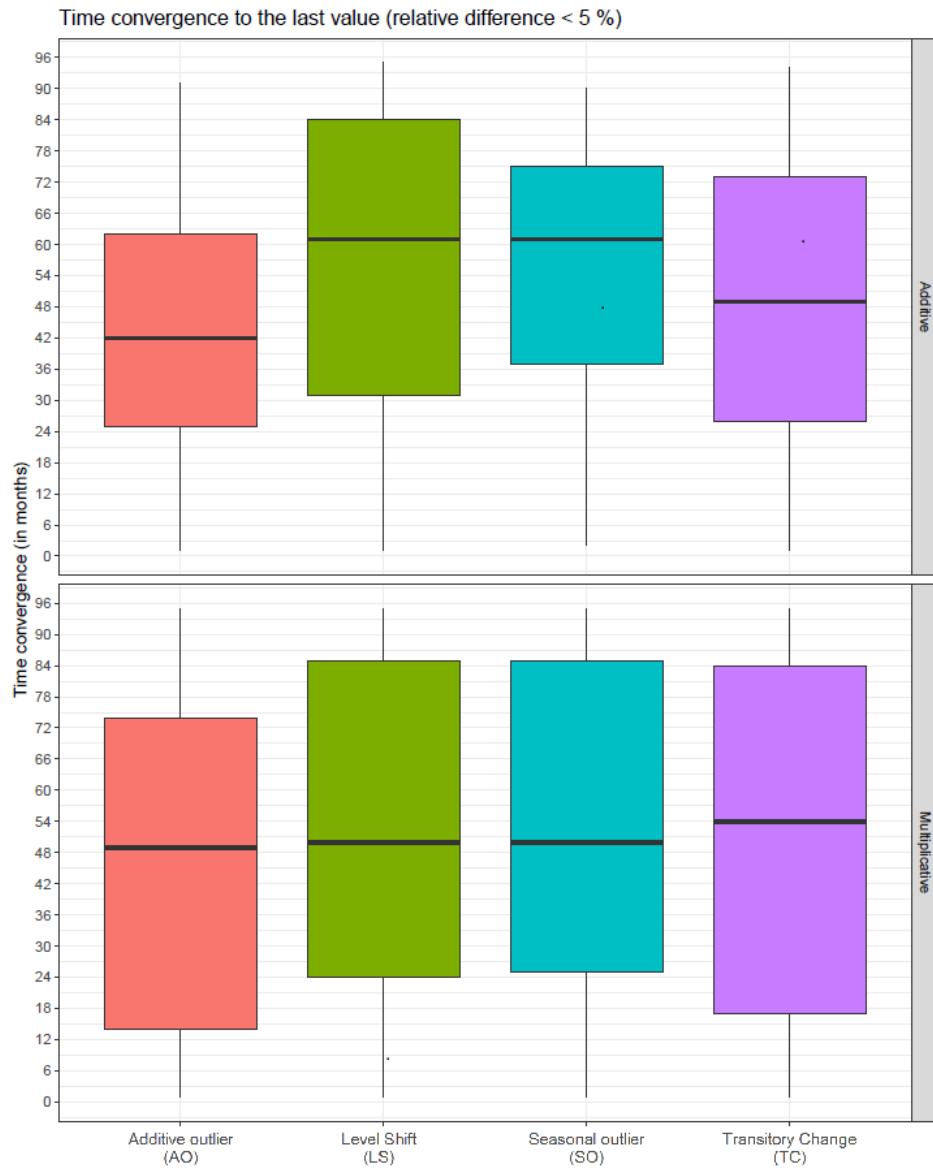
Figure 11: Time convergence (in months) of the coefficient of the break by the decomposition model.
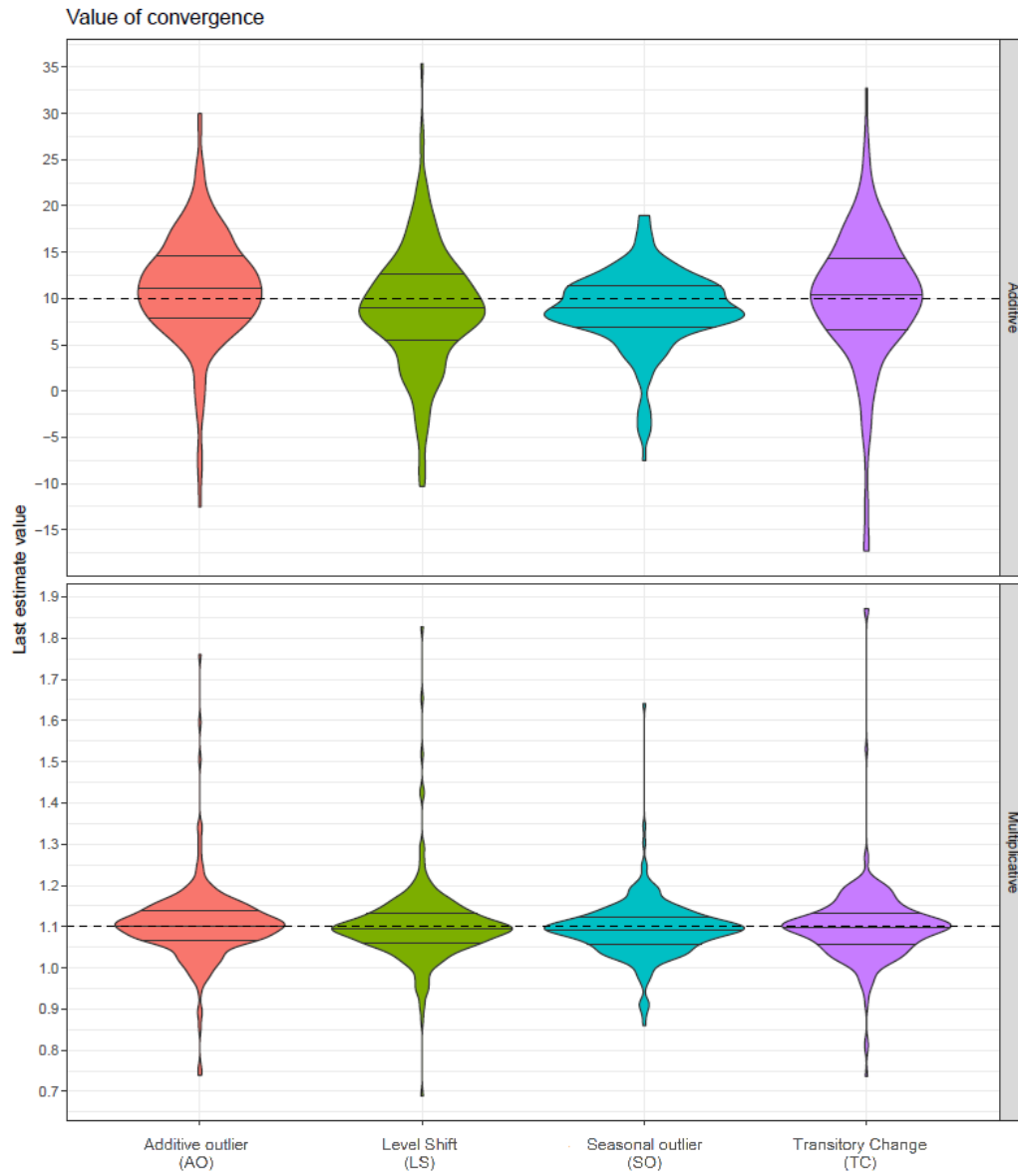
Figure 12: Final estimate of the coefficient of the coefficient of the break by the decomposition model.

Reading note: on the ordinate, the probability density of the estimate coefficient of the studied breaks. The horizontal lines represent the associated quartiles (minimum, first quartile, median, third quartile and maximum).

Table 2: Statistics on the time convergence (in months) of the coefficient of the break by the decomposition model and the type of break.

	Minimum	25 %	50%	75%	Maximum
Additive models					
AO	1	25	42	62	91
LS	1	31	61	84	95
SO	2	37	61	75	90
TC	1	26	49	73	94
Multiplicative models					
AO	1	14	49	74	95
LS	1	24	50	85	95
SO	1	25	50	85	95
TC	1	17	54	84	95

Table 3: Statistics on the convergence values by the decomposition model and the type of break. Expected values are +10 for additive models and 1.1 (+10%) for multiplicative models.

	Minimum	25%	50%	75%	Maximum
Additive models					
AO	-11.6	7.8	11.1	14.2	36.9
LS	-11.4	5.6	9.3	12.7	49.8
SO	-5.8	7.3	8.8	11.0	31.1
TC	-17.4	6.5	10.2	14.1	47.2
Multiplicative models					
AO	0.72	1.07	1.10	1.14	1.88
LS	0.68	1.07	1.10	1.13	1.81
SO	0.86	1.06	1.10	1.12	1.40
TC	0.73	1.06	1.10	1.14	1.87

4 Estimate of the ARIMA model

In the proposed example, the series follows an additive model and the general Reg-ARIMA model is:

$$z_t = \beta_0 LY_t + \sum_{j=1}^{j=6} \beta_j \left(N_{jt} - N_{7t} \right) + \sum_{l=1}^{l=k} \alpha_l O_{lt} + x_t, \tag{4}$$

where O_l ($l = 1, \ldots, k$) correspond to the possible breakd detected by the algorithm. In this formulation we can:

1. Either let the leap year regressors in the trading days set of regressors automatically generated by the algorithm;

2. Or remove the leap year from this set and introduce it as an external regressor.

Both models are perfectly equivalent from the theoretical point of view. But the algorithms, implementation of the theory, are not the theory! Indeed, estimates of the same model under slightly different conditions may yield to significantly different results. Thus, if we experiment on the RF241 series of the French IPI, we obtain (figure 13):

- In the case where the leap year is modeled in the trading days regressors: a significant leap year effect (at the 5% threshold), calendar effects (working days and leap year effect) are significant (at the 5% threshold), 3 breaks and an ARIMA model $(2, 0, 0)(0, 1, 1)$;

- And in the case where the leap year is considered as an external regressor: a significant leap year effect (at the 5% threshold), no working days effect, 8 breaks and an ARIMA model $(0, 1, 1)(0, 1, 1)$.

Fortunately this surprising phenomenon is relatively rare. It is due to the fact that in the second case the model is poorly specified: by not introducing the leap year regressor in the calendar effects, they are considered significantly equals to 0 and are therefore not retained by the algorithm. The result is a different ARIMA model and different breaks.

The leap year regressor is in the calendar effects	The leap year regressor is in the external regressors
Summary	**Summary**
Estimation span: [1-1990 - 11-2016] 323 observations Trading days effects (7 variables) No easter effect 3 detected outliers	Estimation span: [1-1990 - 11-2016] 323 observations No trading days effects No easter effect 8 detected outliers
Final model	**Final model**
Likelihood statistics Number of effective observations = 311 Number of estimated parameters = 15	**Likelihood statistics** Number of effective observations = 310 Number of estimated parameters = 12
Loglikelihood = -931.813018813206 Standard error of the regression (ML estimate) = 4.745365565197271 AIC = 1893.626037626412 AICC = 1895.2531562704798 BIC (corrected for length) = 3.372719903247878	Loglikelihood = -913.7329646425261 Standard error of the regression (ML estimate) = 4.535310915888911 AIC = 1851.4659292850522 AICC = 1852.5164343355573 BIC (corrected for length) = 3.227343071998285
Scores at the solution 0,005221 -0,001104 -0,030472	Scores at the solution -0,003681 -0,017589

Arima model
[(2,0,0)(0,1,1)]

	Coefficients	T-Stat	P[|T| > t]
Phi(1)	-0,5256	-9,46	0,0000
Phi(2)	-0,2878	-5,17	0,0000
BTheta(1)	-0,7913	-20,56	0,0000

Correlation of the estimates

	Phi(1)	Phi(2)	BTheta(1)
Phi(1)	1,0000	-0,7388	-0,0184
Phi(2)	-0,7388	1,0000	0,0489
BTheta(1)	-0,0184	0,0489	1,0000

Leap year

	Coefficients	T-Stat	P[|T| > t]
Leap year	4,3861	2,65	0,0085

Arima model
[(0,1,1)(0,1,1)]

	Coefficients	T-Stat	P[|T| > t]
Theta(1)	-0,5051	-10,08	0,0000
BTheta(1)	-0,7533	-18,80	0,0000

Correlation of the estimates

	Theta(1)	BTheta(1)
Theta(1)	1,0000	0,0280
BTheta(1)	0,0280	1,0000

User variables

	Coefficients	T-Stat	P[|T| > t]
Vars-1.LY	4,5569	2,92	0,0038

Figure 13: Comparison of the models obtained for the RF241 according to whether the leap year regressor is used or not as an external regressor.

Conclusion

Reg-ARIMA models are parametric models that, like many parametric methods, are very sensitive to atypical values and to the respect of the assumptions on which they are based. This lack of robustness has been known for decades and is at the origin of the robust statistics (see for example Tukey (1977) and Huber and Ronchetti (2011)).

To use them well, it is essential to correctly specify **beforehand** the model and to check that the assumptions of the model are well fulfilled. Therefore:

- even if users are fond of long series, it is extremely unlikely that an economic series follows the same ARIMA model for more than 15 years. In particular, the rebasements and changes of nomenclature often lead to retropolize the series with indicators close but with a probable different dynamic. A long series must therefore be seasonally adjusted by sub-periods — see in this respect Eurostat (2015) and Pham and Quartier-la-Tente (2018);

- an Easter effect on series of industrial production is very unlikely, except perhaps for some very precise levels of nomenclature;

- the set of trading days regressors must be based on economic reasons rather than statistical considerations (see the leap year effect example). In this respect, it is interesting to note that the default set of regressors propose by the current methods (6 regressors, one for each day of the week, or 1 regressor representing the 5 days of the week except weekend) are either unstable or inadequate as shown by Attal-Toubert (2012) or Ladiray (2018).

In terms of seasonal adjustment, most of the instabilities mentioned will have a limited effect on the seasonally and trading days adjusted series[6], especially if X-13ARIMA-SEATS is used. But they will undoubtedly change the story in the short term and eventually introduce unpleasant revisions.

The simulations presented in this study are certainly criticizable and improvable but they have the merit of highlighting the potential instability of these Reg-ARIMA models often used as black boxes. However, the automatic algorithms implemented in X-13ARIMA-SEATS and TRAMO-SEATS are important and very useful tools. The best is certainly to use them in accordance with the "European guidelines for seasonal adjustment" published by Eurostat (2015).

Basing, for example, on these models procedures for choosing trading days regressors or more generally for seasonal adjustment parameters, without relying on economic reasoning and knowledge of the field studied, would be a big mistake... which would probably lead to say about you: *"He uses statistics as a drunken man uses lamp posts, for support rather than illumination"*[7]

---

[6]Particularly because the AO, LS and TC are reintroduced into the seasonally adjusted series at end of the process and because calendar effects are usually low.

[7]Quote widely attributed to Andrew Lang (1844-1912), Scottish writer, but the original quote has never been found.

A   Le calendrier grégorien

Le calendrier grégorien est un calendrier solaire, basé sur le mouvement de la Terre autour du Soleil, et pour lequel l'année est une approximation de l'année tropicale, durée que la Terre prend pour aller d'un point fixe, comme un équinoxe ou un solstice, au suivant. Dans ce calendrier, une année normale est faite de 12 mois — janvier, février, mars, avril, mai, juin, juillet, août, septembre, octobre, novembre et décembre — qui contiennent un nombre de jours égal respectivement à 31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30 et 31. Une année normale contient donc 365 jours, ce qui est malheureusement un peu trop court dans la mesure où la Terre met environ 365 jours et 6 heures pour accomplir une révolution autour du soleil. Une meilleure approximation de l'année solaire est obtenue en ajoutant 1 jour à certaines années, les années bissextiles, pour lesquelles le mois de février aura 29 jours. Une année bissextile est une année divisible par 4 mais pas par 100, sauf si elle est divisible par 400. Ainsi, 1900 ne fut pas une année bissextile, 2000 en était une et 2100 ne sera pas bissextile. On a donc *in fine* 97 années bissextiles sur une période de 400 ans. Il reste cependant une petite erreur d'approximation d'environ 1 jour tous les 4000 ans que le calendrier grégorien ne prend pas en compte.

Toute période de 400 ans contient donc $400 \times 365 + 97 = 146097$ jours, soit exactement 20871 semaines. Le calendrier grégorien est donc périodique de période 400 ans. La longueur moyenne d'une année sur ce cycle est de $146097/400 = 365,2425$ jours et la longueur moyenne d'un mois est de $365,2425/12 = 30,436875$ jours.

Bibliography

Attal-Toubert, K. (2012), Régresseurs pour effets de calendrier : comment les construire, comment les choisir ? Actes des 11ᵉᵐᵉˢ Journées de Méthodologie Statistique, `http://jms-insee.fr/jms2012s14_3/`.

Bell, W. R. (1992), Alternative Approaches to Length of Month Adjustment, Research Report Series 1992-17, Satistical Research Division, U. S. Census Bureau, Washington DC. `https://www.census.gov/ts/papers/rr92-17.pdf`.

Eurostat (2015), The ESS guidelines for seasonal adjustment, Eurostat manuals and guidelines, Product Code: KS-GQ-15-001. `http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-15-001`.

Findley, D., McElroy, T. (2018), Background and Perspectives for ARIMA Model-Based Seasonal Adjustment, in Handbook on Seasonal Adjustment, edited by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg. `ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001`.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B.C. (1998), New capabilities and methods of the X-12-ARIMA seasonal adjustment program, Journal of Business and Economic Statistics, 16, 127-152.

Gómez, V. and Maravall, A. (1996), Programs TRAMO and SEATS. Instructions for the User, (with some updates), Working Paper 9628, Servicio de Estudios, Banco de España.

Gómez, V., Maravall, A. (1998), Automatic Modeling Methods For Univariate Series, Working Paper 9808, Servicio de Estudios, Banco de España.

Grudkowska, S. (2017), JDemetra+ Reference Manual Version 2.2, available from URL: `https://ec.europa.eu/eurostat/cros/content/software-jdemetra_en`.

Huber, P. J., Ronchetti, E. M. (2011), Robust Statistics, Wiley Series in Probability and Statistics, John Wiley and Sons.

Ladiray, D., Quenneville, B. (2001), Seasonal Adjustment with the X-11 Method, Lecture Notes in Statistics No 158, Springer, New York. Available in French on `https://www.census.gov/ts/papers/x11_french.pdf`.

Ladiray, D. (2018), Calendar effects, in Handbook on Seasonal Adjustment, edited by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg. `ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001`.

Maravall, A., Caporello, G. (2004), Program TSW: Revised reference manual. Technical Report, Research Department, Bank of Spain.

Maravall, A., López, R., and Pérez, D. (2015), Reliability of the Automatic Identification of ARIMA Models in Program TRAMO, in Beran, J., Feng, Y., and Hebbel, H. (eds.) Empirical Economic and Financial Research. Theory, Methods and Practice, Springer International Publishing, Switzerland, 105-122.

Maravall, A., López, R., and Pérez, D. (2016), Reg-ARIMA Model Identification: Empirical Evidence, Statistica Sinica, 26, 1365-1388.

Maravall, A. (2018), Quality of Seasonal Adjustment in the Model-Based Approach of TRAMO-SEATS, in Handbook on Seasonal Adjustment, edited by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg. `ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001`.

Mehrhoff, J. (2018), Outlier Detection and Correction, in Handbook on Seasonal Adjustment, edited by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg. `ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001`.

Pham, H., Quartier-la-Tente, A. (2018), Désaisonnaliser les séries très longues par sous-période, gains et choix de la longueur de traitement — exemple des séries de l'IPI, Actes des 13èmes Journées de Méthodologie Statistique.

Tukey, J. W. (1977), Exploratory Data Analysis, Addison-Wesley Publishing Company.