Seasonal Adjustment of Daily Data

Dominique Ladiray, Jean Palate, GianLuigi Mazzi and Tommaso Proietti

Paper prepared for the 16th Conference of IAOS
OECD Headquarters, Paris, France, 19-21 September 2018

Session 1.A., Wednesday, 19/09, 11h00: Use of Big Data for compiling statistics

Dominique Ladiray
dominique.ladiray@insee.fr
INSEE

Gian Luigi Mazzi
gianluigi.mazzi@gopa.lu
GOPA Luxembourg

Jean Palate
jean.palate@nbb.be
National Bank of Belgium

Tommaso Proietti
tommaso.proietti@uniroma2.it
University of Rome Tor Vergata

Seasonal Adjustment of Daily Data

DRAFT VERSION 31/08/2018

Prepared for the 16th Conference of the
International Association of Official Statisticians (IAOS)
OECD Headquarters, Paris, France, 19-21 September 2018

Note:
This Paper is a slightly modified version of Chapter 29 of the "Handbook on Seasonal Adjustment" published by Eurostat in May 2018.

ABSTRACT

High frequency data, i.e. data observed at infra-monthly intervals, have been used for decades by statisticians and econometricians in the financial and industrial worlds. Weekly data were already used in the 20's by official statisticians to assess the short-term evolution of the Economy. For example, Crum (1927) studied the series of weekly bank debits outside New York city from 1919 to 1026 and proposed a method to seasonally adjust these data based on the median-link-relative method developed by Persons (1919).

Nowadays, these data are ubiquitous and concern almost all sectors of the Economy. Numerous variables are collected weekly, daily or even hourly, that could bring valuable information to official statisticians in their evaluation of the state and short-term evolution of the Economy. But these data also bring challenges with them: they are very volatiles and show more outliers and breaks; they present multiple and non integer periodicities and their correct modeling implies numerous regressors: calendar effects, outliers, harmonics.

The current statistician's traditional toolbox, methods and algorithms, has been developed mainly for monthly and quarterly series; how should these tools be adapted to handle time series of thousands observations with specific characteristics and dynamics efficiently?

We present some ideas to adapt the main seasonal adjustment methods, and especially "the X11 family" i.e. methods based on moving averages like X11, X11-ARIMA, X12-ARIMA and X-13ARIMA-SEATS. We also make some recommendations about the most appropriate methods for pretreatment and filtering of daily and weekly data.

Keywords: Seasonal adjustment, high-frequency data, ruptures, calendar effects.

Seasonal Adjustment of Daily and Weekly Data

Dominique LADIRAY [1]

August 31, 2018

[1]Institut national de la statistique et des Ã©tudes Ã©conomiques, Département des Méthodes Statistiques, 18 boulevard Adolphe Pinard, 75014 Paris, France. Email: dominique.ladiray@insee.fr

Abstract

Note: Version 0. All comments are welcome.

The annual growth rate or simple moving averages are tools commonly used to comment the evolution of an economic time series. It compares the situation at date t to the one observed a year, 4 quarters or 12 months, before. This very simple tool has quite attractive properties: it is easy to compute and understand, it removes a large part of the seasonality usually present in monthly or quarterly economic indicators, it removes also a large part of the long term trend, and amplifies some of the business cycle frequencies.

But it has also some drawbacks that are often unknown to the analyst and that makes it sometimes difficult, not to say dangerous, to use. In particular, the annual growth rate introduces by design a phase shift and the analyst does not comment the current state of the Economy but the situation in which the Economy was a few months ago. Moreover, the annual growth rate might introduce spurious cycles when applied to a non seasonal series.

This paper proposes an in-depth analysis of the annual growth rate and of the associated annual difference operator. The pros and cons of these tools are demonstrated and illustrated on several main economic indicators.

1 Introduction

High frequency data, i.e. data observed at infra-monthly intervals, have been used for decades by statisticians and econometricians in the financial and industrial worlds. Weekly data were already used in the 20's by official statisticians to assess the short-term evolution of the Economy. For example, Fisher (1923) proposed a weekly index number of wholesale prices that was the first general weekly index number to appear. The index was based on the price quotations of 200 commodities and published in the press each Monday, covering the week ending the previous Friday noon. Crum (1927) studied the series of weekly bank debits outside New York city from 1919 to 1026 and proposed a method to seasonally adjust these data based on the median-link-relative method developed by Persons (1919).

Nowadays, these data are ubiquitous and concern almost all sectors of the Economy. Numerous variables are collected weekly, daily or even hourly, that could bring valuable information to official statisticians in their evaluation of the state and short-term evolution of the Economy.

But these data also bring challenges with them: they are very volatiles and show more outliers and breaks; they present multiple and non integer periodicities and their correct modeling implies numerous regressors: calendar effects, outliers, harmonics.

The current statistician's traditional toolbox, methods and algorithms, has been developed mainly for monthly and quarterly series[1]; how should these tools be adapted to handle time series of thousands observations with specific characteristics and dynamics efficiently?

This chapter presents some ideas to adapt the main seasonal adjustment methods, namely "the X11 family" i.e. methods based on moving averages like X11, X11-ARIMA, X12-ARIMA and X-13ARIMA-SEATS[2] and methods based on Arima models like TRAMO-SEATS[3]. STL, a non-parametric method proposed by Cleveland et al. (1990) and very similar to the "X11 family" in philosophy, is also considered. Unobserved component models are used to complete the set of seasonal adjustment methods. The chapter also makes some recommendations about the most appropriate methods for pretreatment and filtering of daily and weekly data.

Section 2 explains why high frequency data are important for the short term evolution of the Economy and highlights the main characteristics of these data. Section 3 focuses on the usually necessary pretreatment of the data and more specifically on the detection and estimation of outliers and calendar effects. Sections 4 and 5 present adapted methods that can be used to seasonally adjust these data. These methods are illustrated and compared on an example in Section 6 and section 7 concludes.

Not all problems have been resolved at this stage and further research is needed, mainly on the "tuning" of the methods, to assure a good seasonal adjustment of daily and weekly data. For examples, an automatic determination of the length of the filters in non-parametric methods like STL or X11 has still to be found and the use of unobserved component models still requires a difficult selection of relevant harmonics or regressors to obtain a parsimonious model.

2 The Main Characteristics of High Frequency Data

2.1 Why to use High Frequency Data?

Even if daily data have been available for a long time, for examples in the financial and industrial domains, they have not been really incorporated in the indicators constructed to analyze the short-term evolution of the economy. Nowadays, the "Big Data" phenomena makes numerous high frequency data available at a small costs and official statisticians seriously consider them.

[1] With some notable exceptions like the method used by the Bureau of Labor Statistics to seasonally adjust weekly time series, see Cleveland et al. (2018); and of course unobserved component models that can handle any kind of periodicity by design, see Harvey (1989) and Harvey (2018).

[2] See Ladiray and Quenneville (2001), Ladiray (2018) and Findley et al. (1998).

[3] See Gómez and Maravall (1997) and Maravall (2018)

Data	Period (number of observations per cycle) Minute	Hour	Day	Week	Month	Quarter	Year
Annual							1
Quarterly							4
Monthly						3	12
Weekly					4.348125	13.044375	52.1775
Daily				7	30.436875	91.310625	365.2425
Hourly			24	168	730.485	2191.455	8765.82
Half-hourly			48	336	1460.97	4382.91	17531.64
Minutes		60	1440	10080	43829.1	131487.3	525949.2
Seconds	60	3600	86400	604800	2629746	7889238	31556952

Table 1: Periodicity of the series according to the data collection rhythm.

Some of these data might for example be used:

- To improve the timeliness of usual indicators: could we use some high frequency data to get an estimate of the quarterly GDP 15 days after the end of the quarter, or even before?
- To get higher frequency estimates of important variables: a weekly indicator of job vacancies, of consumer price index, a monthly GDP etc.
- To construct new indicators such as climate or sentiment indexes.

It has to be noted that if you want to use these data to build a rapid estimate, you have to remove the components which have also been removed in the target monthly and quarterly series, and only these components. In particular, if the objective if to get a rapid estimate of the seasonally adjusted quarterly GDP, the high frequency data used should be corrected from any calendar and periodic effects.

2.2 Characteristics of Daily and Weekly Data

Figure 1 presents the consumption of electricity in France recorded from January 1st, 1996 to April 30th, 2016: the top graph shows the daily figures and the bottom graph the monthly values. The series shows a clear periodic effect linked to the seasons: the electricity consumption is higher in Winter than in Summer. But there is also a “hidden” weekly effect due to the fact many factories and businesses are closed during the week-end which explains a lower consumption on Sundays and Saturdays.

2.2.1 Multiple and Non-integer Periodicities

Daily data usually present several “regularities”. In many economic series, the 7 days of the week have different behaviors. In Christian countries, many businesses and retail shops are closed on Sunday which directly impacts the turnover of the concerned economic sectors. Apart this “week cycle” of period 7 days, daily series can show an intra-monthly effect of average period 30.436875 days: for example many wages and revenues are paid at the end of the month and this will have an effect on the sales, the money in circulation and other economic indicators. We also have the “solar cycle” mentioned before that is likely to be present through a “year cycle” of average period 365.2425 days.

Table 1 shows the potential periods that could be present in the data according to the data collection rhythm. For example, a hourly series might have up to 5 periodic components: a “Day cycle” (24 hours), a “Week cycle” (168 hours), a “Month cycle” (average length of 730.485 hours), a “Quarter cycle” (average length of 2191.455 hours), and a “Year cycle” (average length of 8765.82 hours).

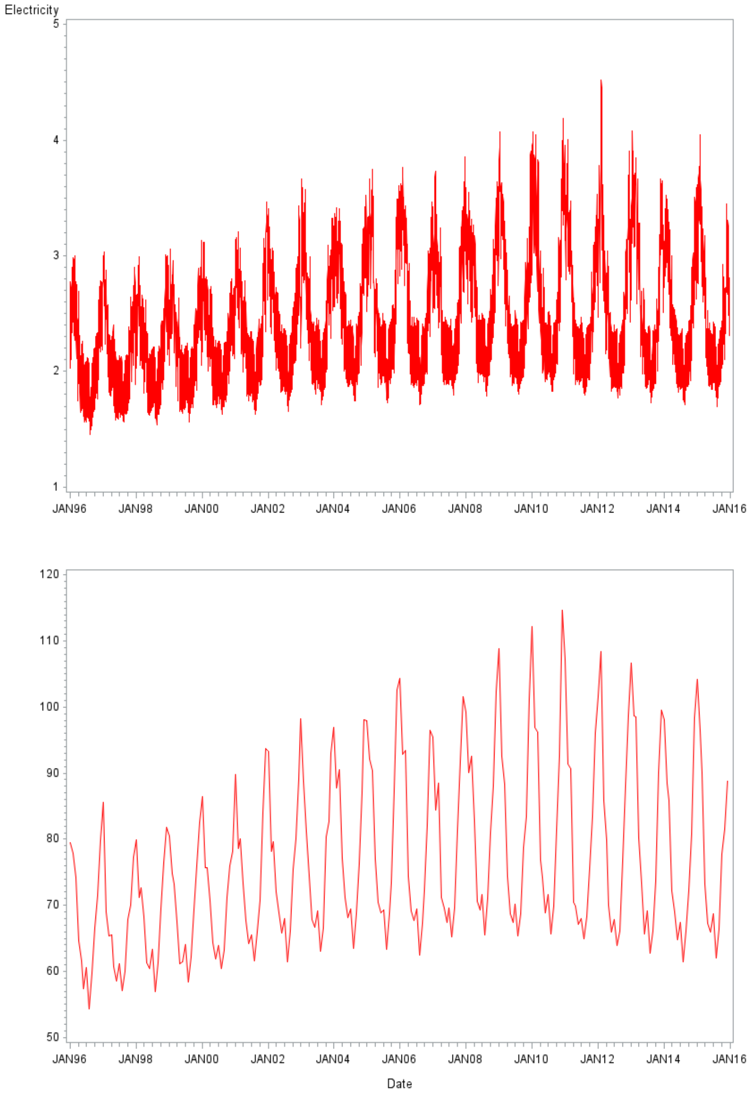

Figure 1: Consumption of electricity in France since 1996: the daily values are in the upper panel, the monthly values in the lower panel.

2.2.2 Important Remarks

1. Hopefully, for daily data at least, the various possible periodicities even if they can be non-integer, are co-primes and it should be possible to separate the various periodic components. The problem might be more complex for hourly and other higher frequency data.

2. More generally, an important problem of the seasonal adjustment of high frequency data will be the proper identification of the various effects and the distinction between the component frequencies: between the various periodic components but also between the trend-cycle frequencies and the annual frequencies.

3. The high number of observations usually available with high frequency data is somehow misleading. For example, if you have 2 years of daily data, you have 730 observations but the estimation of the annual effect might be tricky even if this effect is evident as in the Electricity series. To illustrate the problem, remember that in non-parametric methods like STL or X11, a periodic effect is estimated by smoothing the values corresponding to a modality of the period. Thus, the weekly effect will be obtained smoothing 7 different series: one for each day of the week (Monday values, Tuesday values etc.). This should not be a problem in our 2-year series as we have about 100 observations for each day. But, when it comes to the annual effect, things are more complicated as you just have 2 observations for each day of the year.

   It turns out that the span of the series, the number of annual cycles, is very important and in fact guides the modelling of the series and the quality of the results.

2.3 Checking for the Various Periodicities in the Data

The detection of the various periodicities must be done before any modelling of the time series. Among the statistical tools that can be used in this respect, the most efficient are certainly: the spectrum of the series, the Ljung-Box test and the Canova-Hansen test. Of course, these tools must be adapted to the characteristics of high frequency data.

2.3.1 Spectral Analysis

Spectral analysis is commonly used to check for the presence of seasonality in monthly and quarterly series. But in this context, they do not seem to be very efficient as they might miss a periodicity and have a quite high “false alarm rate” as shown in Lytras et al. (2007).

But, in the context of high-frequency data, the large number of observations improves the quality of the spectrum estimate and might make this tool more efficient.

Figure 2 represents the Tukey spectrum of the Electricity series (upper panel); the periodic behaviour shown in Figure 1 is partly explained by a weekly periodic component: the peaks that can be observed are located at frequency $2\pi/7$ and its harmonics.

The annual periodicity we are expecting in these data is in fact hidden by the 7-day periodic component. The 365-day frequency is associated with fundamental frequency $2\pi/365$ ($360/365 = 0.9863$ degrees) and its harmonics. If we focus on the low frequencies, as shown in the lower panel of Figure 2, we can see the expected but small peaks at the annual frequencies.

2.3.2 Statistical Tests

Ljung-Box Seasonality Test

The Ljung-Box seasonality test checks for the presence or absence of auto-correlation at the seasonal lags. For a time series potentially presenting a seasonality of order k, the test statistic

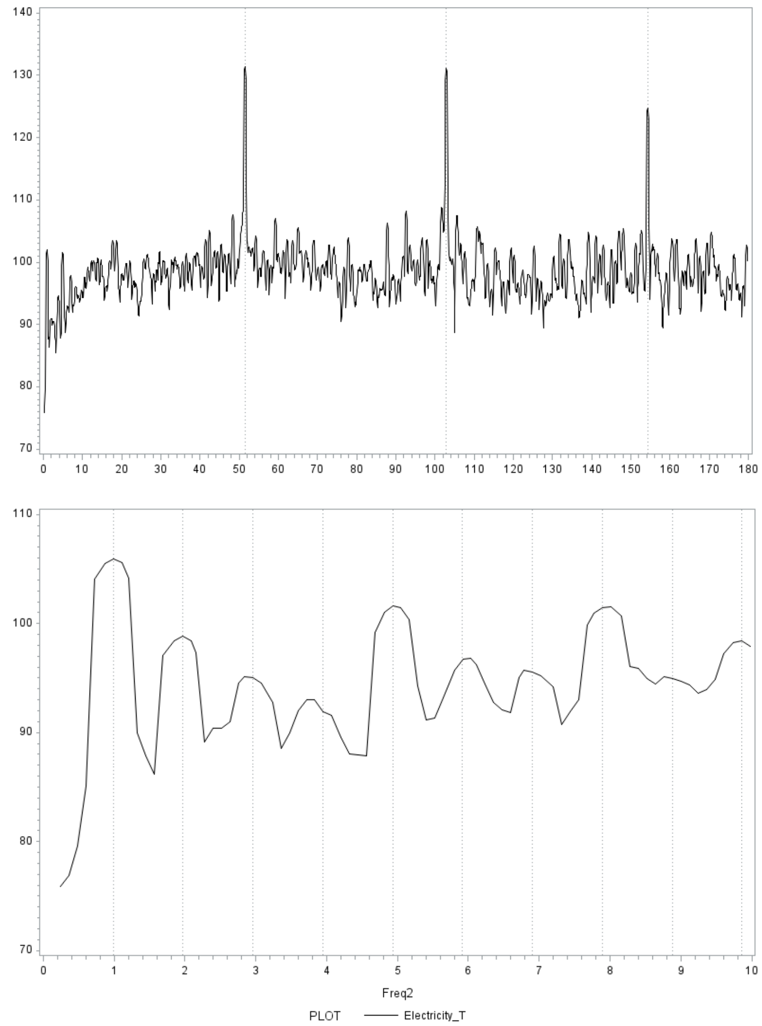

Figure 2: Tukey's Spectrum of the Electricity series (upper panel) and focus on the low frequencies (lower panel).

Period P	Ljung-Box	LB Pvalue	QS stat	QS Pvalue	Autocorrelation at lag			
					P	2*P	3*P	4*P
7	7314.9749	0.000	7314.9749	0.000	0.63502	0.47239	0.43973	0.46221
365	1128.6499	0.000	1128.6499	0.000	0.23905	0.19045	0.15697	0.13547

Table 2: Autocorrelations and Ljung-Box statistics

is:

$$LB = n(n+2)\sum_{j=1}^{h} \frac{\rho_{j*k}^2}{(n-j*k)}$$

where n is the number of observations, ρ_{j*k} is the autocorrelation at lag $j * k$, and h is the number of lags being tested. Under H_0 the statistic LB follows a chi-squared distribution with h degrees of freedom. Usual implementations of the Ljung-Box seasonality test focus on the 2 first seasonal lags and QS follows a $x^2(2)$. Thus, $QS > 5.99$ and $QS > 9.71$ would suggest rejecting the null hypothesis at 95% and 99% significance levels, respectively.

Maravall (2012) suggests that a significant LB statistic can be used to conclude that a seasonality is present only when the sign of the autocorrelation coefficients is consistent with such a hypothesis. He introduces the QS statistics, derived from the Ljung-Bix statistics where, for monthly data, negative values of ρ_1 and ρ_{24} are replaced by zero, and $QS = 0$ for $\rho_{12} <= 0$. This refinement implies that the presence of seasonality will be detected if there is a statistically significant positive autocorrelation at lag 12 or if there are a non negative sample autocorrelation at lag 12 and a statistically significant positive autocorrelation at lag 24.

Ljung-Box seasonality tests can be adapted to daily and weekly data, after proper differencing of the data:

- You can test all the possible relevant periods. For example, for weekly data, you compute the test using lags (52, 53, 104, 106, ...). For daily data, it would be on lags (7, 14, 21, ...) for the weekly periodicity and on lags (365, 366, 730, 732, ...) for the yearly periodicity.
- You can restrict the test to the closest integer. For weekly data, the average yearly periodicity is 52.18, so you use lags (52, 104, ...). For daily data, the weekly periodicity is 7 so you still use lags (7, 14, 21, ...) and as the average length of the year is 365.2475 days, you use for the yearly periodicity lags (365, 730, ...).

Table 2 presents the autocorrelations and associated Ljung-Box statistics for the Electricity series. The results confirm the presence of a significant effect at the weekly and annual periodicities.

Canova-Hansen Test

The stochastic nature as well as the presence of seasonality can be subjected to formal statistical tests. Canova and Hansen (2003) and Busetti and Harvey (2003) derive the locally best invariant test of the null that there is no seasonality against a permanent seasonal component, that can be either deterministic or stochastic, or both.

Consider a seasonal cycle with period p and angular frequency $\lambda = 2\pi/p$. The seasonal component is decomposed into a deterministic term, arising as a linear combination with fixed coefficients of sines and cosines defined at the frequency λ, plus a nonstationary stochastic term, which is a linear combination of the same elements with random coefficients:

$$\gamma_t = \gamma_t^D + \gamma_t^S.$$

Defining $z_t = [\cos\lambda t, \sin\lambda t]'$, $\gamma_t^D = z_t'\gamma_0 + a_0$, where γ_0 and a_0 are fixed coefficients. The stochastic component is $\gamma_t^S = z_t' \sum_{i=1}^{t} k_i$ where k_t is a bivariate vector of serially independent disturbances with zero mean and covariance matrix $\sigma^2 W$.

The null hypothesis is then formulated as $H_0 : \gamma_0 = 0, \sigma^2 = 0$; a permanent seasonal component is present under the two alternatives: $H_a : \gamma_0 \neq 0, \sigma^2 = 0$ (deterministic seasonality),
$H_b : \gamma_0 = 0, \sigma^2 > 0$ (stochastic seasonality).

The test statistic proposed by Busetti and Harvey (2003) is consistent against both alternative hypotheses, and it is computed as follows:

$$w = \frac{1}{n^2\sigma^2} \sum_{t=1}^{T} \left[\left(\sum_{i=1}^{t} e_i \cos \lambda i \right)^2 + \left(\sum_{i=1}^{t} e_i \sin \lambda i \right)^2 \right], \tag{1}$$

where e_i are the OLS residuals obtained from the regression of y_t on a set of explanatory variables accounting for a constant or a linear deterministic trend. Under the null w is asymptotically distributed according to a Cramer von Mises (CvM) distribution with 2 degrees of freedom.

For the frequencies $\lambda = 0$ and $\lambda = \pi$ the test takes the form

$$w = \frac{1}{n^2\sigma^2} \sum_{t=1}^{T} \left(\sum_{i=1}^{t} e_i \cos \lambda i \right)^2, \tag{2}$$

and the null distribution is Cramer von Mises (CvM) with 1 degree of freedom. When $\lambda = 0$, the test statistic is the usual KPSS test of stationarity at the long–run frequency.

The test of the null that the seasonal component is deterministic (H_a) against the alternative that it evolves according to a nonstationary seasonal process (H_b), i.e. characterised by the presence of unit roots at the seasonal frequencies ω_j, is based on the Canova-Hansen (CH) test statistic which is (1), with e_i replaced by the OLS residuals that are obtained by including the trigonometric functions in z_t as additional explanatory variables along with x_t. Under the null of deterministic seasonality, the test has again a Cramér von Mises distribution.

The test statistic in (1) requires an estimate of σ^2. A nonparametric estimate is obtained by rescaling by 2π the estimate of the spectrum of the sequence e_t at the frequency λ, using a Bartlett window.

Figure 3 displays the values of the CH test for the daily Electricity series, versus the period p of the periodic component. Significant values are detected at the weekly and annual frequencies only, for which the null is rejected.

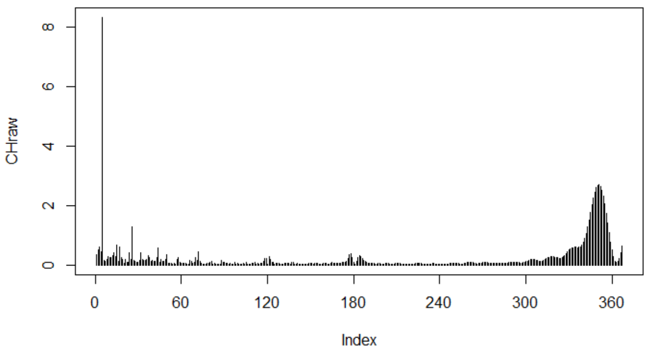

Figure 3: Consumption of electricity in France. Canova-Hansen tests for components with period $p = 2, 3, \ldots, 366$.

3 Pretreatment of Daily and Weekly Data

The seasonal adjustment of high frequency time series poses several challenges. One is the detection of outliers that, if they are not exhibited and imputed, could hamper the proper estimation of the seasonal component. The robustness can be enforced either by an outlier detection procedure or by robust filtering methods. Another is the detection and correction of trading-day effects and moving-holiday effects that are directly linked to the calendar and make periods (quarter, month, week, day etc.) not directly comparable.

Usual seasonal adjustment packages propose an automatic detection facility to detect and correct for outliers and calendar effects. These statistical algorithms have usually be designed specifically for monthly and quarterly time series and are applied before the decomposition of the time series in trend-cycle, seasonal component and irregular. The basic process adopted by both TRAMO-SEATS and X-13ARIMA-SEATS is summarized in Figure 4.

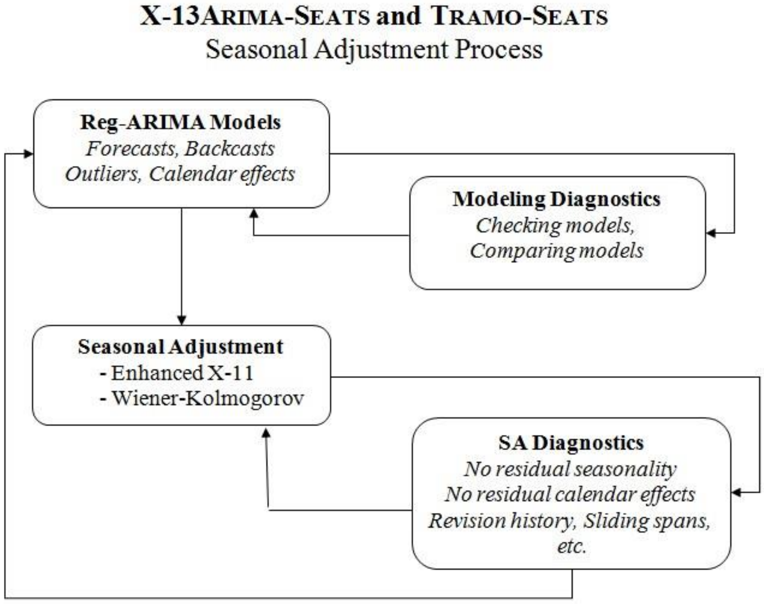

Figure 4: TRAMO-SEATS and X-13ARIMA-SEATS 2-step process.

A common approach is based on Reg-Arima models where each effect (outliers, calendar effects) is associated to a specific regression and where the residuals of the model are supposed to follow a general Arima model. The algorithm proposed by Gómez and Maravall (1998), which is implemented in the software TRAMO-SEATS, is probably the most popular and the most efficient. This algorithm performs an automatic detection of the decomposition model (additive, multiplicative), an automatic detection and correction of outliers (additive outlier, level shifts, transitory changes, ramps, seasonal outliers), an automatic detection and correction of usual trading-day effects, an automatic adjustment of the Arima model and produces forecasts and backcasts of the series.

TRAMO fits a seasonal Arima model $(p, d, q)(P, D, Q)_s$ to a monthly or quarterly series y_t. This model can be written: $\Phi(B)\delta(B)y(t) = \theta(B)E(t)$, where:

$$\delta(B) = \Delta^d \Delta_s^D = (1 - B)^d (1 - B^s)^D$$
$$\Phi(B) = (1 + \varphi_1 \cdots + \varphi_p B^p)(1 + \Phi_1 B^s + \cdots + \Phi_P B^{Ps})$$
$$\theta(B) = (1 + \theta_1 B + \cdots + \theta_q B^q)(1 + \Theta_1 B^s + \cdots + \Theta_Q B^{Qs})$$

In these equations, s is the integer periodicity of the series. The seasonal difference operator is noted $\Delta_s y_t = y_t - y_{t-s}$ and Δ_1 is simply denoted Δ.

For daily and weekly data, s is not any more a single and constant integer. To interpret the peculiar trait of the seasonality, the class of seasonal fractionally integrated ARMA processes, or ARFIMA in short, has been advocated in various contribution, among which Koopman et al. (2007). Following this approach, it is possible to extend the model:

1. Focusing on the "average" periodicities: for a daily series, the possible periodicities will therefore be $s_1 = 7$, $s_2 = 30.436875$ and $s_3 = 365.2425$; for a weekly series, the average yearly periodicity is $s = 52.1775$.

2. Focusing in a first step on a generalization of the airline model with fractional periodicities. The basic models are here:
   - For a daily series (D1):
     $(1-B)(1-B^7)(1-B^{30.436875})(1-B^{365.2425})y_t = (1+\theta_1 B)(1+\theta_2 B^7)(1+\theta_3 B^{30.436875})(1+\theta_4 B^{365.2425})E(t)$
   - For a weekly series (W1):
     $(1-B)(1-B^{52.1775})y_t = (1+\theta_1 B)(1+\theta_2 B^{52.1775})E(t)$

Let us note $s+\alpha$ the periodicity (for example 365.2475 for daily data or 52.1775 for weekly data) where s is an integer (365 or 52) and α a real number belonging to the interval $]0,1[$ (0.2475 or 0.1175).

Using the Taylor expansion of x^α, we have:

$$x^\alpha = 1 + \alpha(x-1) + \frac{\alpha(\alpha+1)}{2!}(x-1)^2 + + \frac{\alpha(\alpha+1)(\alpha+2)}{3!}(x-1)^3 + \cdots$$

And, if we limit to the first two terms of the expansion, we have $x^\alpha \cong (1-\alpha) + \alpha x$. So, we can define the "tilde difference operator":

$$\tilde{\Delta}_{s+\alpha} y_t = y_t - B^{s+\alpha} y_t = y_t - B^s B^\alpha y_t \cong y_t - (1-\alpha)B^s y_t + \alpha B^{s+1} y_t$$

The previous high-frequency models can therefore been rewritten:

1. For weekly series:
   W2: $\Delta\tilde{\Delta}_{52.1775} y_t = (1-\theta_1 B)(1-0.8225\theta_2 B^{52} - 0.1775\theta_2 B^{53})E_t$
2. For daily series:
   D2: $\Delta\Delta_7\tilde{\Delta}_{30.436875}\tilde{\Delta}_{365.2475} y_t = (1-\theta_1 B)(1-\theta_2 B^7)(1-0.563125\theta_3 B^{30} - 0.436875\theta_3 B^{31})(1-0.7525\theta_4 B^{365} - 0.2475\theta_4 B^{366})E_t$
3. And for infra-daily series, similar extensions could be considered with additional periodicities.

3.1 Decomposition Model

Following the genuine TRAMO algorithm it is possible to choose between an additive and multiplicative decomposition model by adjusting the fractional airline model to the raw data and to the log-transformed data and comparing the quality of the adjustments.

But high frequency data are usually very volatile and the presence of outliers might biased the choice towards a multiplicative decomposition. Moreover, because of the presence of several periodicities, the data could present more complex decomposition models: additive for one periodicity, multiplicative for another etc.

To handle these issues, more robust tests like the "spread versus level" test proposed by Hoaglin et al. (1983), Box-Cox transformations might be more appropriate. In this case, robust tests like the "ladder of powers" proposed by Tukey (1977) could also be implemented.

3.2 Calendar Effects

The trading-day effect was originally defined for monthly, quarterly and even annual series. The fact is that months (or quarters) are not directly comparable. They do not have the same number of days (mainly a seasonal effect) and the day composition of months varies from one month to another and from one year to another. For example, May 2015 had 5 Saturdays, one more than May 2014, April 2015 and June 2015. In the retail trade sector, this extra Saturday can make more difficult the year-to-year and month-to-month turnover comparisons. This effect directly linked to the day composition of the month (or quarter) defines the trading-day effect.

Therefore, weekly and daily data cannot theoretically present any pure trading-day effect: weeks always have one day of each kind and the differences between days that could be observed in daily data is a periodic/seasonal component. But it is without taking into account the specificity of the National calendar.

National holidays are often linked to a date, not to a specific day. For example, in catholic countries, Christmas is always the 25^{th} of December, but not always a Sunday. As these National days are usually off, they might impact the activity of some sectors of the economy in different ways. For a weekly (monthly, quarterly) time series, a National holiday contributes to the trading-day effect as the weeks (months, quarters) do not have any more the same number of working days. And this contribution might be significant for a given week.

The way National holidays should be considered might change according to the periodicity of the series. For a daily series, the "Christmas effect" on December 25^{th} and around is a periodic/seasonal effect and should be directly taken into account in the seasonal component. But for a weekly series, this is different as Christmas might impact different weeks and even be in different weeks (the 51^{st} or the 52^{nd}) according to the year. This effect must be very carefully modeled and eliminated. In a similar way, the impact of moving holidays like Easter and other related events, Ramadan, Chinese new year etc. should be specifically modeled.

It is quite easy to derive general impact models[4]. For a given event, the parameters of a model are:

- The date of the event in the Gregorian calendar;
- The span of time on which the event is supposed to have an impact. It can be determined by 4 parameters: the starting date of the impact before the event, the ending date of the impact before the event, the starting date of the impact after the event and the ending date of the impact after the event;
- The nature of the impact: constant, linear, quadratic, increasing, decreasing etc.

Figure 5 and Figure 6 illustrate a general constant impact model and a general "increase-decrease" linear model. The constant impact model supposes the event occurs in period (week) m and has an impact on 8 days in the past, starting 2 days before the event, and on 5 days in the future, starting 3 days after the event. The regressor MH which will be used in the model is, for a given period, proportional to the number of days of impact in this period. Therefore, MH$(m-1) = 4/13$, MH$(m) = 4/13$, MH$(m+1) = 5/13$ and MH$(w) = 0$ for all other periods.

The modeling should also take into account the fact that moving holidays often present a periodical behavior in the Gregorian calendar. For example, the catholic Easter always falls in March or April and more often in April than in March, the orthodox Easter always falls in April or May and more often in April than in May etc. This annual effect should be removed from the regressor that must take into account non-periodical effects.

Calendar Effects in the Electricity series

France celebrates 11 National holidays per year: January 1^{st}, May 1^{st}, May 8^{th}, July 14^{th}, August 15^{th}, November 1^{st}, November 11^{th}, December 25^{th}, Easter Monday, Pentecost Monday and Ascension Thursday.

Just to explore the possible calendar effects, a regressor was created for each of the 11 holidays and for each of the 7 days before and the 7 days after the holiday. We therefore have $11 * 15 = 165$ such regressors. The significance (t-value) of each regressor is reported in Table 3 and we can draw important conclusions for the modelling of these calendar effects.

[4]For a detailed presentation of these models, see Ladiray (2018).

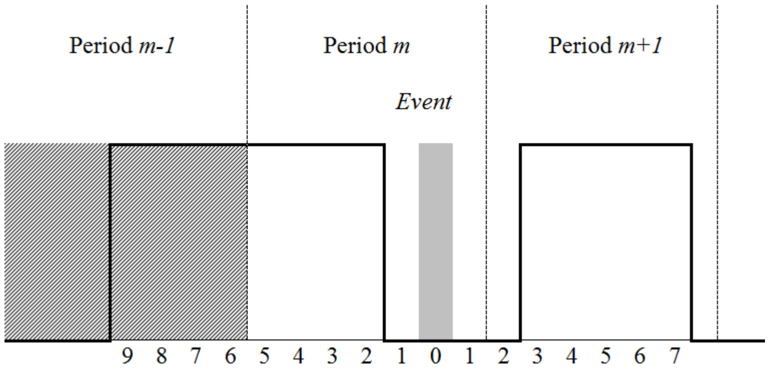

Figure 5: A general constant impact model

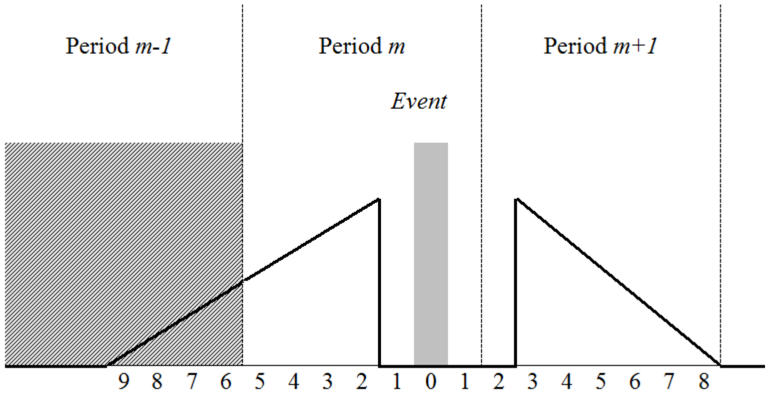

Figure 6: a general "increase-decrease" linear impact model

- As expected, all holidays have a negative effect on the consumption of electricity. But the effects are different which advocates for specific regressors.
- The effect of Pentecost and July 14th is concentrated on the day itself (Lag 0);
- Other holidays show have an impact on the day itself and on the following day;
- The week between Christmas and New Year has a very specific behaviour.

This simple example shows that a proper modelling of a daily series requires to take into account the specificities of the National calendar. Of course, in this case, only the effect of moving holidays (Easter Monday, Ascension and Pentecost) should be taken into account as the effect of the other holidays will be captured by the annual periodic component.

When we restrict the Calendar effect to these 3 regressors, the t-values are:

- -20.128 for the Easter Monday
- -20.963 for Pentecost
- -18.200 for Ascension.

3.3 Dealing with outliers

Both parametric methods (Reg-Arima models or unobserved component models) and non parametric models (X11 and STL) use automatic procedures to detect and correct series for outliers.

These procedures are quite efficient and can be used as they are for daily and weekly data.

- Reg-Arima methods, like TRAMO-SEATS, and unobserved component models use intervention analysis and stepwise iterations;

Table 3: National holiday effects in the daily consumption of electricity in France

Lag	Jan1	May1	May8	Jul14	Aug15	Nov1	Nov11	Dec25	EasterMonday	Ascension	Pentecost
-7	3.581	-1.553	-0.279	0.163	-1.183	-0.013	-0.92	-0.711	0.197	1.401	0
-6	-4.933	-1.313	2.422	0.565	-0.698	-0.382	0.317	-0.822	-0.089	1.228	0
-5	1.013	-1.194	-0.177	1.141	-1.07	-0.385	-0.542	-1.294	1.594	0.8	0
-4	-1.506	0.194	-0.135	1.649	-1.336	-0.364	-2.063	-1.581	1.255	0	0
-3	-0.196	0.611	0.728	1.462	-1.769	-0.145	0.081	-1.759	0.006	1.757	3.86
-2	-0.977	1.953	1.809	1.669		-0.306	0.183	-5.742	1.038	2.25	0.782
-1	-1.095	-0.168	0.419	-0.595		-2.943	-1.81	-13.247	0	2.729	0
0	-5.782	-6.881	-6.644	-17.457	[illegible]	-16.265	-15.455	-12.638	-20.188	-17.219	-18.976
1	-7.971	-3.412	-2.562	-1.092	-4.903	-4.18	-2.712	-1.309	-2.151	-7.614	1.262
2	-1.429	1.198	0.626	2.692	-1.682	-0.2	-0.832	-4.863	-0.438	-1.53	3.262
3	0.356	0.839	0.98	2.121	-1.406	0.497	-0.097	-1.569	0.598	3.732	3.45
4	0.178	0.03	0.615	2.129	-1.069	-0.275	0.139	-2.352	1.308	0	3.899
5	-0.437	-1.276	0.419	1.88	-1.217	0.968	0.14	-1.966	0.481	4.091	0.329
6	-0.65	-0.412	0.632	1.452	-0.903	2.069	0.111	-3.858	0.949	3.978	1.726
7	-1.001	1.116	1.519	1.795	-0.361	-0.25	-0.634	-1.181	0	4.354	0

Table 4: Outliers in the daily consumption of electricity in France

Type	Date	Estimate	Pvalue
AO	29/12/1996	0.152	4.399
AO	08/05/1997	0.184	5.066
AO	01/05/2008	0.202	5.548
AO	20/12/2009	0.155	4.5
AO	30/12/2009	-0.192	-5.5
AO	31/12/2009	-0.205	-5.858
AO	25/12/2010	0.182	5.138
AO	12/02/2012	0.155	4.499

- X11 has a robust built-in procedure based on trimmed means to detect and correct for additive outliers that is usually preceded by a Reg-Arima correction procedure;
- STL has a robust built-in procedure based on LOESS, a local weighted regression procedure that can be also complemented if needed by a Reg-Arima correction procedure.

Outliers in the Electricity series

As shown by Table 4, only 8 outliers are detected, a very small number considering the length and the volatility of the series. This is a sign that the modeling strategy fits quite well the data.

Moreover, most of the detected outliers (6 out of 8) concern the Christmas/New Year week or National holidays for specific years. Once again these results advocate for a precise modeling of the National holidays.

4 Seasonal Adjustment Based on Non-parametric Methods

X11 and STL share a common non parametric philosophy. The 2 methods are based on an iterative estimation of the series components. The difference lies in the choice of the filters: a set of moving averages for X11, and the use of local weighted regressions (LOESS) for STL. Both software can be extended to daily and weekly data.

4.1 Adapting X11 to multiple and non integer periodicities

The iterative process of X11 is precisely described in Ladiray and Quenneville (2001) and seems to be easily extended to multiple periodicities.

4.1.1 An X11-like algorithm for multiple periodicities

Let us suppose our time series X_t presents 2 different seasonalities of order p and q and can be represented by the following additive model:

$$X_t = TC_t + S_{p,t} + S_{q,t} + I_t.$$

A basic algorithm to derive an estimation of the various components, and therefore of the seasonally adjusted series, can be summarized in 10 steps.

1. **Estimation of the trend-cycle by a composite moving average**

$$TC_t^{(1)} = M_{p\times q}(X_t)$$

The symmetric moving average used here is a so-called $p \times q$ moving average, of order $(p+q)-1$, which preserves linear trends, eliminates order-p and order-q constant seasonalities and minimizes the variance of the irregular component. Its coefficients are the coefficients of the product of the 2 polynomials $\frac{1}{p}(1+t+t^2+\cdots+t^p)$ and $\frac{1}{q}(1+t+t^2+\cdots+t^q)$ [5].

2. **Estimation of the global seasonal-irregular component**

A first estimation of the global seasonal-irregular component can be easily obtained removing the trend-cycle estimation from the raw data.

$$(S_{p,t} + S_{q,t} + I_t)^{(1)} = X_t - TC_t^{(1)}$$

3. **Estimation of each basic seasonal-irregular component**

Smoothing the global seasonal-irregular estimate with a simple moving-average of order p, which removes order-p constant seasonalities, gives an estimate of a seasonal-irregular component corresponding to periodicity q.

$$(S_{q,t} + I_t)^{(1)} = M_p\left[(S_{p,t} + S_{q,t} + I_t)^{(1)}\right]$$

And, by difference, we obtain an estimate of the seasonal-irregular component corresponding to periodicity p.

$$(S_{p,t} + I_t)^{(1)} = (S_{p,t} + S_{q,t} + I_t)^{(1)} - (S_{q,t} + I_t)^{(1)}$$

Note that it is better to begin with the smallest periodicity in order to minimize the number of lost points at the ends of the series.

4. **Estimation of each basic seasonal component, by a 9-term Henderson moving average over each period**

For seasonality p, we smooth each of the p series corresponding to each period with a 9-term Henderson moving average which preserves local quadratic trends. The seasonal factors are then normalized so that their sum over each p-period span is approximately zero.

$$(S_{p,t})^{(1)} = H_9\left[(S_{p,t} + I_t)^{(1)}\right]$$

and

$$\left(\tilde{S}_{p,t}\right)^{(1)} = (S_{p,t})^{(1)} - M_p\,(S_{p,t})^{(1)}$$

[5] If the order (p+q)-1 is even, we use a $2 \times p \times q$ moving average.

We do the same for seasonality q.

$$(S_{q,t})^{(1)} = H_9\left[(S_{q,t} + I_t)^{(1)}\right]$$

and

$$\left(\tilde{S}_{q,t}\right)^{(1)} = (S_{q,t})^{(1)} - M_q\,(S_{q,t})^{(1)}$$

5. **Estimation of the irregular component, detection and correction of the extreme values**

   A first estimate of the irregular component is derived, removing the seasonal factors from the estimate of the global seasonal-irregular component.

$$I_t^{(1)} = (S_{p,t} + S_{q,t} + I_t)^{(1)} - \left(\tilde{S}_{p,t}\right)^{(1)} - \left(\tilde{S}_{q,t}\right)^{(1)}$$

   A robust moving standard error σ_t, the F-pseudosigma [6], is computed for each data t. Each value of the irregular component I_t is assigned a weight ω_t, function of the standard deviation associated with that value, calculated as follows:

   - Values which are more than $2.5\sigma_t$ away in absolute value from the average 0 are assigned zero weight.
   - Values which are less than $1.5\sigma_t$ away in absolute value from the average 0 are assigned a weight equal to 1.
   - Values which lie between $1.5\sigma_t$ and $2.5\sigma_t$ in absolute value from the average 0 are assigned a weight that varies linearly between 0 and 1, depending on their position.

   and the irregular is therefore corrected:

$$\tilde{I}_t^{(1)} = \omega_t \times I_t^{(1)}$$

6. **Estimation of the global seasonal-irregular component corrected from extreme values**

   A new estimation of the global seasonal-irregular component, corrected from extreme values is then derived:

$$(S_{p,t} + S_{q,t} + I_t)^{(\bullet)} = \left(\tilde{S}_{p,t}\right)^{(1)} + \left(\tilde{S}_{q,t}\right)^{(1)} + \tilde{I}_t^{(1)}$$

7. **Estimation of each basic seasonal-irregular component**

   Smoothing the global seasonal-irregular component corrected from extreme values with a simple moving-average of order p gives a new estimate of a seasonal-irregular component corresponding to periodicity q.

$$(S_{q,t} + I_t)^{(2)} = M_p\left[(S_{p,t} + S_{q,t} + I_t)^{(\bullet)}\right]$$

   And, by difference, we obtain an estimate of the seasonal-irregular component corresponding to periodicity p.

$$(S_{p,t} + I_t)^{(2)} = (S_{p,t} + S_{q,t} + I_t)^{(\bullet)} - (S_{q,t} + I_t)^{(2)}$$

[6] The F-pseudosigma is defined as $\sigma_t = \frac{IQR}{1.349}$ where IQR is the interquartile range.

8. **Estimation of each basic seasonal component, by a 9-term Henderson moving average over each period**

   For seasonality p, we smooth each of the p series corresponding to each period with a 9-term Henderson moving average. The seasonal factors are then normalized so that their sum over each p-period span is approximately zero.

$$(S_{p,t})^{(2)} = H_9\left[(S_{p,t} + I_t)^{(2)}\right]$$

   and

$$(\tilde{S}_{p,t})^{(2)} = (S_{p,t})^{(2)} - M_p\,(S_{p,t})^{(2)}$$

   We do the same for seasonality q.

$$(S_{q,t})^{(2)} = H_9\left[(S_{q,t} + I_t)^{(2)}\right]$$

   and

$$(\tilde{S}_{q,t})^{(2)} = (S_{q,t})^{(2)} - M_q\,(S_{q,t})^{(2)}$$

9. **Preliminary estimation of the seasonally adjusted series**

   A first estimate of the seasonally adjusted series can be easily computed by removing the seasonal factors from the raw data:

$$A_t^{(1)} = X_t - (S_{p,t})^{(2)} - (S_{q,t})^{(2)}$$

Remark

Another simple idea to extend the genuine X11 algorithm to multiple periodicity is to use it successively to remove each seasonality. For a daily series, you use the genuine algorithm to remove first the periodicity 7 after choosing the relevant filters. The intra-monthly and annual periodicities will then be part of the trend-cycle component. Then, you use the adjusted series to remove the intra-monthly periodicity etc.

4.1.2 Dealing with non integer periodicities

To illustrate the problem, let us suppose that our daily series presents an intra-monthly seasonality which means that the days of the month (first, second, third etc.) are not similar. For example, the pay-day (every 2 weeks, end of the month etc.) might have an impact on the retail sales.

To estimate the effect of each day of the month, X11 will smooth a "ragged matrix" similar to the one presented in Table 5 which presents 2 problems:

- The first one, due to a length-of-month effect, is the presence of missing values in the last columns of the matrix which makes the smoothing by moving averages impossible.
- The second problem is that, if you assume a strong end-of month effect, the values for the 27th day of the month are not completely coherent as they might be affected by the end-of-month effect for February and not for the other months.

Therefore, a correct imputation of missing values depends of the nature of the effect you want to estimate. You can either interpolate the missing values per column or use some kind of "time wrapping" for each line putting the last values of the month in the last columns and interpolating the missing values.

The annual periodicity presents the same problem linked to the Leap Year as the matrix to smooth will have missing values in its 366th columns. But, in this case, it is suggested to ignore the problem by smoothing only the 365 columns and interpolating the missing values in the annual seasonal component.

Date	D01	D02	D03	D04	....	D26	D27	D28	D29	D30	D31
12/2011	0.28	0.40	0.34	0.38	....	0.42	0.34	0.29	0.32	0.30	0.35
01/2012	0.29	0.20	0.09	0.02	....	-0.51	-0.45	-0.48	-0.45	-0.41	-1.21
02/2012	-1.17	-0.29	-0.28	-0.26	....	-0.58	-0.58	-0.57	-0.57	.	.
03/2012	-0.47	-0.40	-0.44	-0.37	....	-0.78	-0.77	-0.73	-0.62	-0.41	-0.45
04/2012	-0.47	-0.41	-0.41	-0.24	....	-0.50	-0.30	-0.30	-0.27	-0.22	.
05/2012	-0.23	-0.21	-0.15	-0.16	....	-0.04	-0.10	-0.07	-0.14	-0.05	0.03
06/2012	0.11	0.10	0.06	0.02	....	-0.26	-0.22	-0.05	0.11	0.15	.
07/2012	0.08	0.08	0.05	0.04	....	0.18	0.27	0.26	0.20	0.26	0.28
08/2012	0.35	0.45	0.48	0.48	....	0.27	0.15	0.13	0.17	0.39	0.49
09/2012	0.47	0.47	0.48	0.38	....	0.32	0.48	0.61	0.63	0.62	.
10/2012	0.73	0.77	0.77	0.78	....	0.34	0.35	0.37	0.39	0.43	0.49
11/2012	0.51	0.59	0.61	0.56	....	-0.16	-0.16	-0.09	0.11	0.23	.
12/2012	0.23	0.24	0.20	0.20	....	0.97	1.01	1.12	1.12	1.09	1.06
01/2013	1.09	0.94	0.77	0.68	....	-0.49	-0.51	-0.53	-0.52	-0.45	-0.27
02/2013	-0.22	-0.24	-0.25	-0.29	....	-0.74	-0.68	-0.54	.	.	.
03/2013	-0.43	-0.49	-0.50	-0.54	....	-0.64	-0.40	-0.07	-0.07	-0.09	-0.09

Table 5: An example of “ragged matrix” encountered in the estimation of an intra-monthly periodicity

4.2 Improving STL for multiple periodicities

STL[7] and X11 share the same iterative and nonparametric philosophy . . . and face the same problems. The idea of iteratively adjusting several seasonal movements in STL has already been put forward by Cleveland et al. (1990); and a generalization of STL to multiple periodicities has been proposed by Ollech (2016).

In the case of daily time series, the main obstacle is the constraint of STL to have a constant period length, i.e. the number of observations per period has to be constant. This is not unique to STL though, as we already saw X11 has the same requirements. While this is unproblematic in the case of intra-weekly seasonality, the number of days per month and the number of days per year are not identical for all periods. In any case, the period length has to be standardized, either by omitting a subset of the data or by artificial prolongation.

Before decomposing the series and estimating the various seasonal components with STL, the series may have to be pre-adjusted for calendar effects, outliers and missing values. The proposed generalized Reg-ARIMA solution could be use for this purpose. If the time series has not been outlier adjusted, the robust version of STL should be used, i.e. the number of iterations of the outer loop needs to be higher than 0.

The iterative procedure proposed by Ollech is the following:

- **Step 1: Removing Intra-Weekly Seasonality** For most of the LOESS regressions in STL there exist sensible default values for the degree of smoothing (see Cleveland et al. (1990)), i.e. the number q of neighbourhood points. In the case of intra-weekly seasonality, q has to be large enough so that the week-day effects do not get confounded by other effects such as moving holidays.
- **Step 2: Removing Intra-Monthly Seasonality** While the length of any year is always the same except for one day, the number of days in a given month is quite irregular ranging from 28 to 31 days with an average of 30.4 days per month. STL is flexible enough to handle a time series with a frequency of 30.4, but this estimation strategy will blend together the effects of different dates.

  As in X11, the problem of confused effects can be circumvented by extending each month to cover 31 days. For the extension of the time series, several different methods could be used. While regression based techniques have the advantage of possibly incorporating much or all of the information that is contained in the actual series, in the context of high frequency time series they can become computationally burdensome. Therefore, Ollech

[7] See Cleveland et al. (1990) for a detailed presentation of the method.

suggests to use a simpler approach, namely cubic splines. Their benefit is an extremely fast algorithmic implementation and a high degree of smoothness, which seems preferable to a simple linear interpolation. A Forsythe-Malcolm-Moler algorithm is for example applied to obtain values for the additional data points needed to have 31 days in each month. The only parameter that needs to be specified is the value for q, which determines the length of the seasonal filter.

- **Step 3: Removing Intra-Annual Seasonality**
  As the time series has been extended in the pre-treatment step, the excess days including every 29^{th} of February have to be removed so that each year contains 365 days. Then q has to be chosen. As outliers and the influence of moving holidays have already been removed from the series, the main objective in choosing q is the variability of the intra-monthly seasonality.

  Yet, the robust version of STL may be chosen to ensure that intra-annual seasonal peaks do not affect the estimation of the intra-monthly seasonal factors.
- **Step 4: Producing the Seasonally Adjusted Series**
  The 29^{th} of February is added back to the seasonally adjusted series via spline interpolation. Finally, the effects of the outliers are added to the original and seasonally adjusted time series.

4.3 Pros and cons of the X11 and STL approaches

The 2 approaches are computationally very efficient and they can easily handle long and complex high frequency time series.

Taking into account non integer periodicities requires the imputation of missing values but solutions have been proposed that seem efficient. The real problems are mainly in the choice of the filters and in the tuning of their length according to the component to estimate.

- From this point of view, STL is slightly simpler as it essentially uses the LOESS robust regression. The number of points q to take into account in the regression must anyway be precised.
- X11 is certainly more complex to tune as the moving averages might be different according to the component you want to estimate. In the genuine X11 algorithm, and for monthly and quarterly series, the order of the moving averages is selected according to a "signal to noise" ratio (the I/C ratio for the order of the trend cycle moving averages and the I/S ratio for the seasonal moving averages). Large scale simulations have still to be done to understand the behavior of these ratios and to elaborate a decision rule for high frequency data.

5 Seasonal adjustment based on parametric models

5.1 Seasonal adjustment based on ARIMA Models

The TRAMO algorithm has been adapted to account for multiple periodicities and adapted to daily and weekly data (see Section **??**). The challenge is more difficult for SEATS as the current algorithm cannot handle high degree AR and MA polynomials.

5.2 Canonical decomposition

For weekly models, the usual canonical decomposition of SEATS applies but for daily models, we have to introduce a additional models corresponding to the intra-monthly and weekly periodicities. The different components are derived using the auto-regressive polynomials precised in Table 6.

Table 6: Auto-Regressive Polynomials used by the new implementation of SEATS

Component	Auto-regressive polynomial
Trend	$(1-B)^3$
Weekly	$1+B+B^2+\cdots+B^6$
Intra-monthly	$(1+B+\cdots+B^{30})$ or $(1+B+\cdots+B^{29}+0.436875B^{30})$
Annual	$(1+B+\cdots+B^{364})$ or $(1+B+\cdots+B^{364}+0.2475B^{365})$
Noise	1

It has to be noted that 365, 30 and 7 are relatively prime numbers, so that the decomposition is well defined.

5.3 Estimation of the components

Taking into account the size of the problem, the estimation of the components corresponding to the canonical decomposition is tricky. The Burman's algorithm can be used for the weekly models provided that the moving average coefficients are not too close to 1 (we can of course force them to be lower than a given value, for instance 0.95, as it is done in the current SEATS). The main advantages of that algorithm is that it is very fast, that it uses few resources and that it does not imply the computation of the actual models of the canonical decomposition (the pseudo-spectra are sufficient). However, the Burman's algorithm can become rapidly unstable especially in the case of daily models.

For such models, we recommend to use the Koopman's disturbance smoother. Unfortunately, this algorithm yields other numerical challenges:

- The factorization of very large polynomials (to get the actual models of the components);
- The diffuse initialization of complex state space forms;
- The stabilization of the disturbance smoother itself.

These challenges have been solved by using some non-standard solutions:

- Polynomial reduction by least squares techniques;
- Diffuse squared root initialization;
- Small correction mechanism in the disturbance smoother.

5.4 Pros and Cons of the current implementation

1. The "TRAMO-SEATS" algorithm applied to daily and weekly data works quite well, is fast and reliable. It is based on a clear methodology.

2. It gives in particular a fast and sound way to clean the raw data from outliers and trading-day effects.

3. At the moment, the adapted program is based on a generalization of the airline model to multiple and non-integer periodicities only. It is likely that this model will not suit all series. In particular, the effect of the differencing on the annual frequencies should be investigated more precisely.

4. The stability of the model should also be checked very carefully.

5.5 Seasonal adjustment based on unobserved components models

The unobserved components (UC) approach provides a flexible and easily customizable approach to modelling and adjusting daily and weekly time series. For the series analysed in the course of the project, a "basic" specification, encompassing most of the stylized features of the series, is the following:

$$y_t = \mu_t + \sum_{j=1}^{m} \gamma_{jt} + \beta' x_t + \epsilon_t. \tag{3}$$

- The trend component, μ_t, is usually specified as a local linear trend, as in the monthly case. The trend model is already flexible enough, featuring two sources of disturbances, one for the level and one for the rate of change (velocity, or slope) of the trend.
- The seasonal component, $\sum_{j=1}^{m} \gamma_{jt}$, for daily data, consists of the annual cycle and its harmonics, the intra-monthly cycle and the weekly cycle. Two representations for this component can be considered, the trigonometric representation and the periodic spline representation. These are interchangeable in regular cases, but offer comparative different advantages in nonstandard cases.
- The regression component accounts for the effect of moving festivals, such as Easter, and other calendar components. It can be used also for modelling special seasonal effects, such as the Christmas effect.
- The irregular component, ϵ_t, is assumed to be uncorrelated.

5.5.1 Pros of the UC approach

As stated above, the main comparative advantage of the UC approach, over the reduced form ARIMA approach and the nonparametric one, is the flexibility by which special features can be incorporated. Nothing prevents, in principle, to model the trend using a cubic spline with fixed knots, or a segmented trend, subject to slope changes, rather than having a model where the level and the velocity are subject to small and frequent shocks. Also, the definition of the components is well delineated and the identifiability of the various effects is imposed a priori. Furthermore, the variability of the components is learnt from the data, as the model is estimated by maximum likelihood. As a result the filter used for seasonal adjustment is not fixed, but it is adapted to the specific nature of the series under investigation.

The seasonal component can be represented by the combination of trigonometric cycles, which are defined at the seasonal frequencies. As the cosine and sine functions at the harmonic and the fundamental frequencies are orthogonal, model selection is simplified, and the decision on the number of trigonometric terms to be included can be carried out by the usual information criteria. Hence, the trigonometric specification makes variable selection relatively easy, due to the orthogonality of the seasonal cycles.

An alternative effective representation for the seasonal component is the periodic spline model. Here the model selection problem deals with locating the knots. In many application, a quite natural choice is locating the knots at the end of each month. The advantage of the periodic spline model, over the trigonometric one, is the possibility of entertaining special periods of the year, e.g. Christmas, by allocating a specific knot. This however comes with a disadvantage, as the effect associated to a new knot is not orthogonal to the other effects already present. As a result, model selection is more complicated.

Finally, it has to be stated that the UC approach can handle missing values and can easily incorporate a complete pretreatment of the series, with the purpose of adjusting for outliers and structural breaks.

5.5.2 Limitations of the UC approach

In the actual implementation of the UC approach we can envisage two classes of limitations: those related to the estimation of the model and those related to the specification of the model.

The model depends on a set of hyperparameters, typically a transformation of the variance parameters of the disturbances driving the evolution of the latent components, which have to be estimated by maximum likelihood. Usually, the optimization takes place by a quasi-Newton algorithm, such as BFGS, using numerical first and second derivatives, which is iterated until convergence. The following problems may arise:

1. The maximization algorithm requires initial parameter values. It is not trivial, in general, to provide good starting values of the hyperparameters.
2. When the number of observations is very large, and or the state vector is high-dimensional, each iteration of the algorithm may take a substantial computational time.
3. The evaluation of the likelihood for certain hyperparameter values may fail.
4. There is always the possibility that several local maxima are present and that the algorithm converges to a local maximum.

One problem with the specification of the unobserved components is that the models tend to be very “series specific”, i.e. rather ad hoc. For instance, knots selection can be rather arbitrary; we may or may not have a knot corresponding to the 25^{th} of December; or we may add an intervention variable for the Christmas effect, which spreads the effects over neighbouring days and repeats itself constantly over the years.

As a consequence, it appears that it is difficult to implement a general user friendly software that can be applied to a vast number of situations. The user could get the impression that modelling daily and weekly time series is an art, rather than a routine operation.

A specification issue we often face is deciding a priori whether a particular component is deterministic or stochastic. This actually is a problem when the frequency of the observations is high, then the permanent of trend shocks (to the component μ_t) will tend to be small, so that a time evolutive trend may be actually very difficult to estimate. As a matter of fact, it is indeed very difficult to distill what is permanent from a daily observation: a needle in a haystack. Hence, the estimates of the size of the disturbances to μ_t tend to pile up at zero.

When the number of years of available data for a daily series is not large, it is a problem to identify the trend from the annual cycle: as a matter of fact, the annual cycle corresponds to the frequency $2\pi/365.25 = 0.0055\pi$, which is very close to the zero frequency. As a result, it will be difficult to isolate the trend from the fundamental annual cycle. Specifying a model with both time evolving trend and annual cycle is asking too much from the data.

6 Application: Seasonal adjustment of the Electricity series

The Electricity series has been decomposed using the 4 methods: X11, STL, TRAMO-SEATS and UCM. As already noticed, the series has 2 different periodicities: a weekly component (Sunday and Tuesday do not have the same behavior) and an annual component (Winter and Summer have different behaviors).

6.1 The seasonal patterns

The weekly patterns estimated by the 4 methods are presented in Figure 7. It clearly appears that the 4 methods deliver a similar message; the weekly pattern is very stable. We observe a decreasing of the consumption of electricity during the weekend (-5% on Saturday and -12% on Sunday) a slight augmentation on Monday (+1%) and a stable consumption the 4 other days

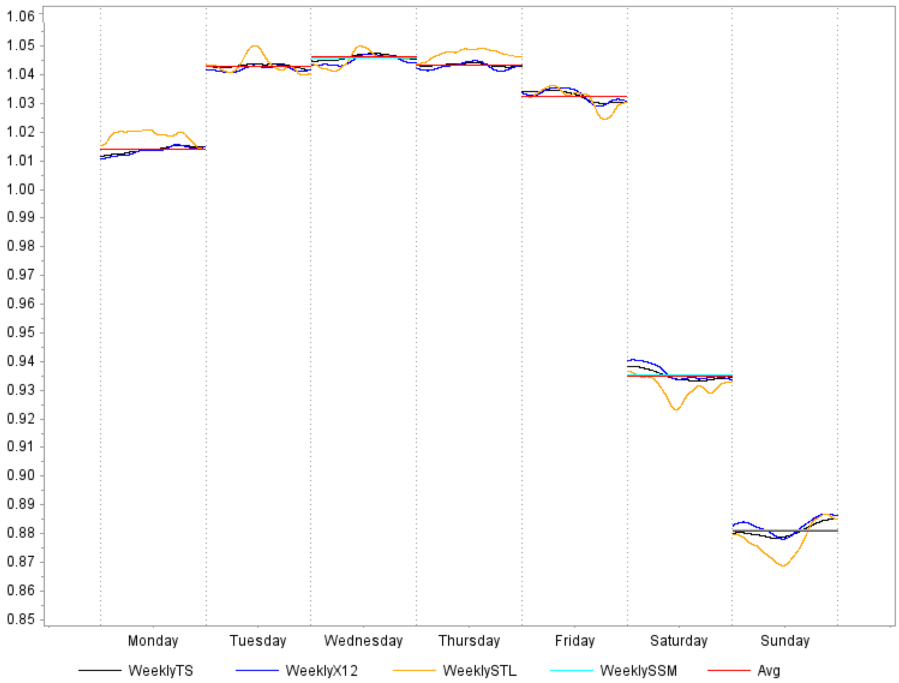

Figure 7: Various estimates of the weekly pattern in the Electricity series
The patterns obtained with X11, STL, TRAMO-SEATS and UCM are represented respectively in black, blue, orange and cyan. The red line represents the average of the TRAMO-SEATS weekly pattern.

(about +4%). This common weekly pattern is clearly linked to the fact that businesses and factories are closed during the weekend when retail sale shops are mainly closed on Sunday and Monday.

On the opposite the yearly patterns shown in Figure 8, if they deliver roughly the same message, are quite different:

- All patterns present a maximum in the Winter period (January-February) with a consumption of electricity about 30% above the average and a minimum in Summer about 15% under the average. The patterns also show 2 specific periods where the consumption of electricity is very low: the week between Christmas and New year period (only 10% above the average) and the big period of holidays in August (more than 25% under the average).
- Yearly patterns are supposed to reflect the different behavior of each day of the year. In particular, National holidays should appear as specific as most of the businesses, shops and factories are closed. From this perspective, the UCM yearly pattern appears too smooth.

6.2 Validation of the results

The validation of the results, the quality of the decomposition, must first respect some basic ideas. In particular, the "seasonally adjusted" series, as well as the irregular series, should present no residual seasonality and no residual calendar effect. This can be assessed in particular using the statistical tests or graphs previously presented: Ljung-Box seasonality tests, Canova-Hansen test, spectra etc.

The results of the Canova-Hansen tests performed on the 4 estimates of the seasonally adjusted series are presented in Figure 9.

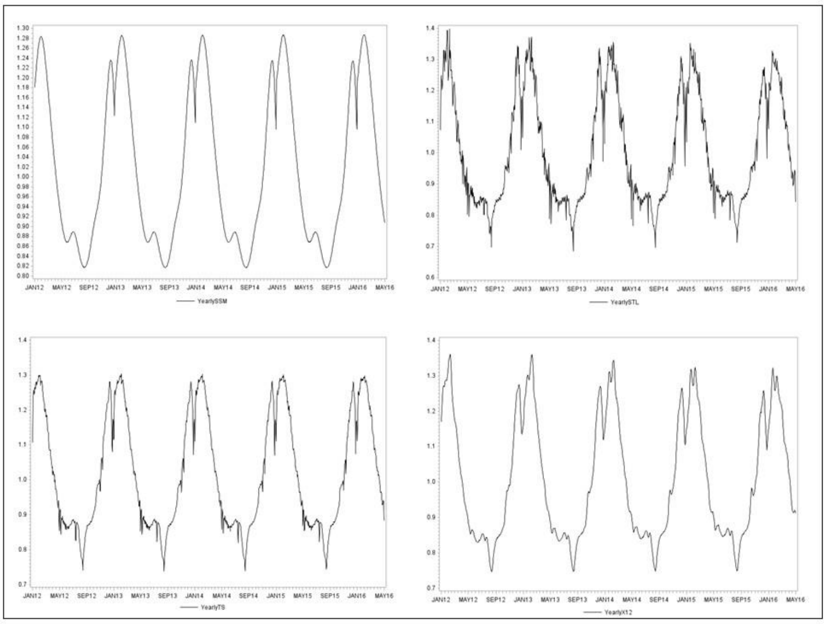

Figure 8: Various estimates of the yearly pattern in the Electricity series
These patterns were obtained with UCM (upper left panel), STL (upper right panel), TRAMO-SEATS (lower left panel) and X11 (upper right panel)

Table 7: Tests on the presence of residual seasonality in the 4 seasonally adjusted series

		X11		UCM		TS		STL	
Test	Period	Value	Pvalue	Value	Pvalue	Value	Pvalue	Value	Pvalue
SFtest	7	17.64	0.00	0.07	0.99	0.35	0.91	62.49	0.00
QStest	7	409.47	0.00	275.28	0.00	375.48	0.00	386.85	0.00
SFtest	365	0.90	0.89	1.19	0.01	0.45	1.00	0.69	1.00
QStest	365	389.18	0.00	473.40	0.00	28.51	0.00	0	1

- No problem of residual weekly periodicity is detected in the 4 seasonally adjusted series;
- On the opposite, a residual yearly periodicity appears in the UCM and in the TRAMO-SEATS seasonally adjusted. series

Fisher tests on dummy seasonals and QS tests have also been performed to check for residual seasonality in the 4 seasonally adjusted series. Results are presented in Table 7.

- As far as the weekly periodicity is concerned, the QS test detects a residual seasonality in all seasonally adjusted series. And the Fisher test validates the absence of residual seasonality in the UCM estimate.
- For the annual periodicity, the Fisher test validates all seasonally adjusted series except the UCM estimate when the QS test finds significant residual seasonality in all series except the STL one.

It turns out that the adjustments are not really satisfactory.

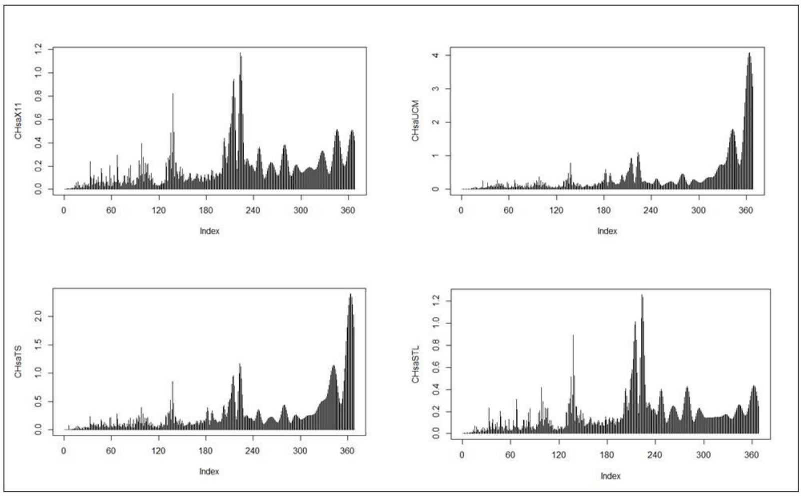

Figure 9: Canova-Hansen test on the 4 seasonally adjusted series of the Electricity series.
These tests concern the estimates obtained with X11 (upper left panel), UCM (upper right panel), TRAMO-SEATS (lower left panel) and STL (upper right panel)

7 Conclusions

This chapter demonstrated that the main seasonal adjustment methods used for monthly and quarterly series, namely TRAMO-SEATS and X12-ARIMA, as well as STL, can be adapted to high frequency data which present multiple and non integer periodicities. TRAMO-SEATS can be modified using fractional Arima models and more efficient numerical algorithms; and the non parametric and iterative processes of X11 and STL can also be easily adapted after imputation of the missing values induced by the different lengths of months and year.

But the tuning of the multiple parameters of the methods might be cumbersome.

- Unobserved component models can by design handle any kind of high frequency data but the models tend to be very "series specific", i.e. rather ad hoc. In particular, the selection of harmonics or knots can be rather arbitrary. As a consequence, it appears that it is difficult to implement a general user friendly software that can be applied to a vast number of situations.
- The tuning of X11 and STL parameters, namely the length of the filters used in the decomposition, needs to be improved. In the genuine X11 algorithm, and for monthly and quarterly series, the order of the moving averages is selected according to a "signal to noise" ratio (the I/C ratio for the order of the trend cycle moving averages and the I/S ratio for the seasonal moving averages). Thresholds have been defined by simulations and practice. Large scale simulations have still to be done to understand the behavior of these ratios and to elaborate a decision rule for high frequency data.
- At the moment, the adapted TRAMO-SEATS program is based on a generalization of the airline model to multiple and non-integer periodicities only. It is likely that this model will not suit all series. The stability of the model should also be checked very carefully.

Despite these shortcomings, it seems clear that methods for the treatment of high frequency time series will improve and be more relevant in the near future, due to the advances in automated data collection and the progress in statistical and econometric analysis.

References

Busetti, F., Harvey, A. (2003). Seasonality tests. Journal of Business and Economic Statistics, 21(3), 420-436.

Canova, F., Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. Journal of Business and Economic Statistics, 13(3), 237-252.

Cleveland, W. P., Evans, T., Scott, S. (2018), Weekly Seasonal Adjustment: A Locally-weighted Regression Approach, in Handbook on Seasonal Adjustment, ed. by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., Terpenning, I. J. (1990), STL: A seasonal-trend decomposition procedure based on Loess (with discussion), Journal of Official Statistics, 1, 6, 3-33.

Crum, W. L. (1927), Weekly Fluctuations in Outside Bank Debits, The Review of Economics and Statistics, 9(1), 30-36.

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, 2, 16, 127-152.

Fisher, Irving (1923), A Weekly Index Number of Wholesale Prices, Journal of the American Statistical Association, Vol. 18, No. 143, pp. 835-840.

Gómez, V., Maravall, A. (1997). Programs TRAMO and Seats : Instructions for the user (beta version: June 1997). Banco de España, Servicio de Estudios, Documento de Trabajo 9628.

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

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press: Cambridge.

Harvey, A.C. (2018), Seasonality and unobserved components models: An overview, in Handbook on Seasonal Adjustment, ed. by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg.

Hoaglin, D., Mosteller, F. and Tukey, J. W. (1983). Understanding Robust and Exploratory Data Analysis, Wiley Series in probability and mathematical statistics, New-York.

Koopman, S., Ooms, M., and Carnero, M. (2007). Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices. Journal of the American Statistical Association, 102(477), 16-27.

Ladiray, D. (2018), Calendar Effects, in Handbook on Seasonal Adjustment, ed. by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg.

Ladiray, D. (2018), Moving Average Based Seasonal Adjustment, in Handbook on Seasonal Adjustment, ed. by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg.

Ladiray, D., Quenneville, B. (2001) Seasonal Adjustment with the X-11 Method, Lecture Notes in Statistics, New York: Springer-Verlag, pp.260.

Lytras, D., Feldpausch, R. and Bell, W. (2007). Determining Seasonality: A Comparison of Diagnostics From X-12-Arima. Paper presented at the Third International Conference on Establishment Surveys (ICES-III), Montréal, Canada.

Maravall, A. (2012). Update of Seasonality Tests and Automatic Model Identification in TRAMO-Seats, Working paper, draft, Bank of Spain.

Maravall, A. (2018). Quality of Seasonal Adjustment in the Model-Based Approach of TRAMO-Seats, in Handbook on Seasonal Adjustment, ed. by G. L. Mazzi, co-edited by D. Ladiray, European Union, Luxembourg.

Ollech, D. (2016). Analysing Daily Time Series, Working Paper, BundesBank.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley, Reading, MA.

Persons, W. M. (1919), Indices of business conditions. The Review of Economics and Statistics, 1, 5-107.