

Discussion Paper

some title or other

The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

2016 | 05

Eugenio Perrucci

Contents

1	Introduction	3
2	Filters in the time domain and the Fourier domain	5
2.1	The Fourier transform	5
2.2	Designing an improved filter	6
2.3	Periodogram	10
3	Seasonal adjustment procedure	11
3.1	JDemetra+	12
3.2	JDemetra+ spectral analysis	12
3.3	TRAMO/SEATS and X-13-ARIMA spectral diagnostics	13
3.4	Squared gain of components filter.	15
4	Filtering a time series with the designed filter	16
4.1	First iteration of the designed filter on a time series	16
4.2	Filtered series as input for JDemetra+ analysis	18
5	Conclusions	21

The identification and removal of the seasonal component is one of the main issues in time series analysis, and it is key operation performed by national statistical agencies such as Statistics Netherlands. Several procedures exist for this goal, available in a wide range of software. The most often applied techniques are based on the use of filters. A linear filter is proposed in this paper, which can be incorporated in existing frameworks and packages for the treatment of time series, and which identifies seasonal influences in one step, with no iteration required. The aim of the study is to evaluate the efficacy of this proposed filter. In particular, the filtered series is used as the input of a seasonal adjustment procedure, performed by JDemetra+. This latter is a software for time series seasonal adjustment recommended by the European Statistical System (ESS) and European System of Central Bank (ESCB). It incorporates two methods: X-13-ARIMA, an enhanced X-11 seasonal adjustment procedure (see (Dagum E.B., 1980)), developed by the U.S. Census Bureau, and TRAMO/SEATS (see (Maravall A., Gomez V., 1997), and (Maravall A., Sanchez J., 2000)), proposed by the Bank of Spain. The performance of the proposed filter is analyzed on a real time series; most of the observed diagnostics focus on the spectral analysis.

1 Introduction

A key concept in time series analysis is the decomposition of a given time series into a trend component, a seasonal component and noise. Seasonality consists in movements repeated throughout the year, with similar intensity in the same period. It means that they are expected to be predictable. Seasonal movements are often large enough such that they can mask other characteristics of the data that are of interest to analysts of current trends. Removal of the seasonal component allows to produce series whose movement are easier to evaluate over time. Indeed, seasonal movements can arise with different timing and amplitude for different time series, making hard their comparison in trend-cycle terms. In this manner, it is also easier to achieve better forecasts. Figure 1.1 shows a decomposition of a time series obtained through a

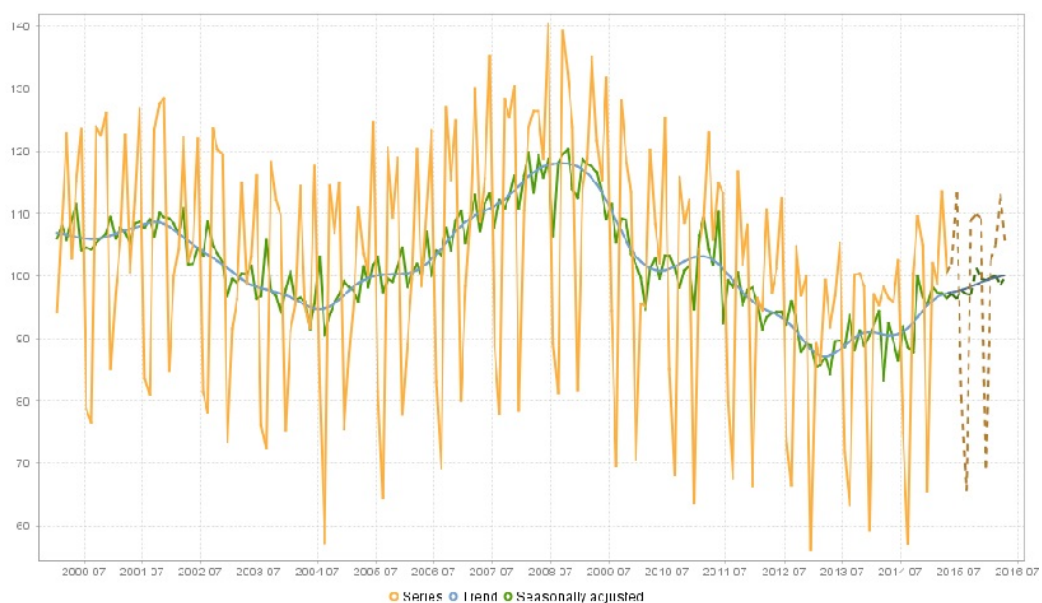


Figure 1.1 A decomposition of the Netherlands volume index of production time series in trend and seasonally adjusted series performed by JDemetra+.

JDemetra+ seasonal adjustment procedure. A common method for obtaining the trend is to use linear filters on given seasonally adjusted series, and in the simplest case moving averages are considered. It means that the filtered value at time t is an average of the values in a neighbour of pre-defined length of that target point. Different seasonal adjustment procedures make use of linear filters. The most often applied filters used to estimate the trend-cycle are the Henderson filters (Henderson, 1916). It is the most widely applied filter to estimate the trend-cycle component in nonparametric seasonal adjustment software, such as the U.S. Bureau of the Census X-12-ARIMA and its variant, X-13-ARIMA. For a good overview of recent developments about capabilities of these software see (Findley D.F., 2005) and references therein. Those procedures are available also in other software, like Gretl, R, EViews, and, particularly, in JDemetra+.

This latter is a statistical software package for seasonal adjustment supported by Eurostat. It includes two methods: X-13-ARIMA and TRAMO/SEATS. Both of the procedures imply a pre-treatment of the time series, aimed to fit an ARIMA model (RegARIMA model) to historical data; after removing the deterministic effects, such as outliers, Easter effects, trading days, leap years etc. etc.), and include those in a regression model (called RegARIMA). The two procedures strongly differ in how the time series decomposition is performed. As mentioned before, X-13-ARIMA works with nonparametric linear filters, while TRAMO/SEATS estimates the various components with signal extraction techniques based on ARIMA models, specified for each component of the series (a detailed discussion is offered by (Maravall A., 2008a)). The estimation of the model parameters is done using the so called Weiner-Kolmogorov filters.

Another relevant topic related to time series analysis is the revisions problem. Many variables, published at sub-annual frequency (monthly or quarterly), produce new data continuously. An effect of this is that the current or most recent observations are subject to revisions as new observations become available. This is because asymmetric filters are usually used to smooth the last observations, and these filters are time-dependent, i.e. there is a different one for each of the last time points. Therefore, as new observations are available, the estimate of the previous observation is modified, because it is re-calculated with a different filter. Then, the seasonally adjusted series need to be updated. Those updates are called revisions. A considerable disadvantage of the available procedures is that revisions produce usually entirely new seasonal adjusted time series. While in practice the updates on older sections of the adjusted time series may be minimal, it is unsatisfactory not to be able to regard any part of the time series as definitive or final. The question is how to reduce the frequency of revisions without losing accuracy in estimation. For a further discussion of the asymmetric filters problem, see (Dagum E.B., Bianconcini S., 2015). The paper is structured as follows. Section 2 describes the proposed filter and discusses its properties in the frequency domain. Section 3 reports the seasonal adjustment procedure carried out on a real time series, characterized by a strong seasonal influence. The analysis is performed using JDemetra+, and it focuses on the spectral analysis of the input and of the output. After adjusting the time series using the proposed filter, in Section 4 the filtered series is considered as the input of JDemetra, to assess whether the removal of the seasonal component obtained by the filter is satisfactory or not.

2 Filters in the time domain and the Fourier domain

For a thorough discussion of the effects that a general filter has on a time series, a valuable diagnostic is the frequency response of that filter. The latter consists in the Fourier transform of the filters factors. The Fourier transform is often also referred to as the Fourier spectrum.

2.1 The Fourier transform

The Fourier transform $Y(\omega)$ of a function of time $y(t)$ is:

$$Y(\omega) \equiv \mathcal{F}(y) = \int e^{i\omega t} y(t) dt \quad (1)$$

with

$$e^{i\omega t} \equiv \cos(\omega t) + i \sin(\omega t)$$

Here ω is the frequency of a signal, related to the period P by:

$$\omega = \frac{2\pi}{P} \quad (2)$$

If the function $y(t)$ is only known at discretely sampled times t_i , which are regularly spaced, a discrete version of the Fourier transform can be defined as well, where the integral is replaced by summation with appropriate weights for each of the samples y_{t_i} , as follows:

$$\mathcal{F}(y) = \sum_t e^{i\omega t} y(t) \quad (3)$$

This paper will refer to y_{t_i} with y_t .

For the Fourier transform there is also an inverse operation.

$$y(t) = \mathcal{F}^{-1}(Y) = \frac{1}{2\pi} \int d\omega e^{-i\omega t} Y(\omega) \quad (4)$$

Since there are particularly efficient algorithms (the Fast Fourier transforms or FFTs) for computing discrete Fourier transforms if the number of samples is exactly a power of 2, normally time series are truncated to an appropriate length or padded out symmetrically with appropriate average values in order to get the number of samples which satisfies this criterion. A useful reference for FFT implementations and their properties is (Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1992).

One property of the Fourier transform, whether continuous or discrete, that will be useful for the further discussion, is that it is a linear operation, and therefore so are the inverse operations. Linearity implies that the Fourier transform of the sum of functions $y(t)$ and $z(t)$ is the sum of their Fourier transforms (Bracewell, 1965):

$$\mathcal{F}(ay + bz) = a \mathcal{F}(y) + b \mathcal{F}(z) \quad (5)$$

where a and b can be any constant.

Of particular importance when carrying out discrete Fourier transforms is a limitation that is imposed by the sampling cadence, that is the time interval between consecutive observations. Indeed, with a finite cadence of sampling, with time interval Δt , it is impossible to detect any periodic signal with a frequency that is higher than the Nyquist frequency. The latter is:

$$\omega_{Nyq} = \frac{\pi}{\Delta t} \quad (6)$$

The Nyquist frequency is related to the number of considered periods T , in the following manner:

$$T = N\Delta_t, \quad \Delta_t = \frac{T}{N}, \quad \text{and so :} \quad \omega_{Nyq} = \frac{\pi N}{T}$$

where N is the number of observations.

A discretely sampled finite time series can be represented without loss of information by its discretely sampled Fourier transform. Since the Fourier transform or spectrum is complex valued, the number of sampling points required is only half the number of points in the time series. If a standard FFT is applied on a time series, these sampling points are uniformly spaced over the frequency interval $[0, \omega_{Nyq}]$.

The reason for discussing Fourier transforms in the context of filtering a time series is that filtering in the time domain corresponds to a particularly simple operation in the frequency domain. In general, filtering in the time domain of a regularly sampled time series can be written in the form of a weighted average:

$$\tilde{y}_t = \sum_{k=-m}^n w_k y_{t+k} \quad (7)$$

where the weights w_k are the filter's factors. This is a general form, allowing for asymmetric filtering. For instance, in causal filters, $n = 0$, so that only information from the previous and present observations of a time point is used and none from the successive ones (this is necessary when estimating time series components in the last time point available). While m and/or n could in principle be infinite, this has no practical purpose in the current context. Also, in the context of seasonal filtering it is more usual to employ symmetric filters so that not only $m = n$ but also $w_{-k} = w_k$. Often an additional property imposed is that $\sum w_k = 1$, but in the applications at hand this is not always used.

The equation (7) above is known as a convolution in the time domain of the function y and the function w , in which the latter is simply the set of averaging weights interpreted as (samples of) a function of time. It can be demonstrated that the Fourier transform of the convolution of two series in the time domain is identical to the product of the Fourier transforms of the two series (Bracewell, 1965). Therefore the Fourier transform of the filtered time series \tilde{y} can be calculated as the product of the Fourier transform of the original time series and the Fourier transform of the filter function w .

$$\mathcal{F}(\tilde{y}) = \mathcal{F}(y * w) = \mathcal{F}(y)\mathcal{F}(w) \quad (8)$$

in which the $*$ indicates the convolution of functions. Multiplication is a much more straightforward operation than calculating the above running averages of (7), and in the frequency domain it is also more straightforward to see the effect of a filter on a signal with a particular frequency or period. In the paper of (Dagum E.B., Chhab N., Chiu K., 1996) the frequency domain properties of many of the filter choices of the X11-ARIMA package are discussed in details, which remain valid also for the most recent X12 implementations. This suggests an alternative approach to create a set of weights, namely by starting with some desirable properties of the filter in the Fourier domain, and then inverse transforming to obtain the appropriate weights w_k .

2.2 Designing an improved filter

Determining the effect that a particular set of weights has in the frequency domain is useful to design a set of weights that is in some sense "optimal". In practical operations, such as those performed by eg. Statistics Netherlands, the intention is to regularly publish or update seasonal adjusted time series and perhaps also the time series of seasonal factors. In this case it may be

useful to apply moving average in the time domain, rather than Fourier transforming the entire time series, carrying out the multiplication in the frequency domain and subsequently inverse transforming again. Useful discussions of the spectral analysis of time domain filters can be found for instance in (Grether D.M., Nerlove M., 1970),(Shumway R.H., Stoffer D.S., 2011) and (Findley D.F., Martin D.E.K., 2006).

In designing the weights it is advisable to have the total extent of non-zero values w as limited as feasible, ie. to make m and n in equation (7) as small as possible.

$$w_k \equiv 0 \quad \forall |k| > \max(m, n). \quad (9)$$

The reason for this is that when updating the published seasonal adjusted time series, only the most recent samples need to be flagged as provisional, whereas further back beyond the m^{th} sampled point in the past the seasonal adjusted series can be considered final, because by design it can no longer change in any regular update.

A separate desirable feature is for the Fourier transform of w to have zero imaginary part, $\Im(W) = 0$, so that the phase of periodic signals is unaffected by the filtering operations. For the Fourier transform of w to be entirely real, the weights w need to be symmetric around 0. For this reason the design will focus on symmetric weights for which:

$$w_{-k} = w_k \quad (10)$$

Another requirement is that the long-term average of the seasonal component has to be equal to 0. Translated to the frequency domain, it means that the value of the Fourier transform of the seasonal component at $\omega = 0$ must be equal to 0. The Fourier transform of the original time series can have any value, so by making use of equation (9) it can be seen that this requirement can be met if the Fourier transform $W(\omega)$ of the weights w for isolating the seasonal component is equal to 0 at $\omega = 0$. Translating the requirement $W(0) = 0$ back to the time domain yields the simple constraint that:

$$\sum_{k=-n}^n w_k \equiv 0 \quad (11)$$

If the weights w are designed to determine the seasonal component, the *trend+ cycle+noise* series can be obtained by subtracting the seasonal component from the original time series. This is the case of the set of filter proposed in here. The linearity property (5) ensures that such additions and subtractions in the time domain are treated identically in the frequency domain. If one has a filter $W(\omega)$ in the frequency domain which is designed to extract seasonal movements from a time series, then the filter $1 - W$ produces a time series with just the *trend+cycle+noise*. This implies that there is an additional desirable property or (soft) constraint for the values W_k in the frequency domain which is that:

$$-\epsilon < W_k < 1 + \delta \quad (12)$$

in which ϵ and δ are as small as feasible given the other constraints. While ideally $\epsilon = 0$ and $\delta = 0$, in practice with a finite set of filter weights w satisfying (10), this may not be achievable. This implies that if the dynamic range of the original time series is very high (ie. the height h_f of a peak in the Fourier transform of the time series is very large) there may still be undesirable remaining seasonal signal in the filtered time series, because δh_f and/or ϵh_f are not sufficiently small.

The standard filtering procedures involved in TRAMO/SEATS and X-13-ARIMA imply an assumption or idealisation that the seasonal movements in a time series are perfectly periodic and repeat identically from one year to the next. If this were indeed the case, then it would be perfectly adequate to design weights that eliminate signal at very specific frequencies that are integer values in the units of cycle/years. The weights of this standard approach, when Fourier

transformed, produce the graph shown in Figure 2.1. It clearly shows notches precisely at integer multiples of 1 cycle/year, broadened because of the finite resolution of the sampling in the frequency domain.

The problem with standard filtering is that signal that is not precisely at these frequencies but in-between, is not included in the seasonal effect. For a real time series this genuinely causes issues because in practice also the “envelope” of the seasonal effect is subject to long term changes. The seasonal pattern may qualitatively be very similar from year to year but most often it will not be completely identical from one year to the next: eg. the amplitude may be well modulated by effects similar to those that determine long-term trends or economic cycles. Such long term effects on seasonal patterns have a signature in the frequency domain: strong peaks at integer multiples of one cycle/year will be substantially broadened, and therefore produce signals at frequencies where the filter corresponding to the naive design of the standard procedures will not include them in the seasonal component and, therefore, also not extract them from the *trend+cycle+noise* series.

For this reason it is more suitable for real time series to design a filter that “removes” all signal in a band between frequencies of around one cycle/year ie. between $\sim 1/12$ cycle/month and $\sim 5/12$ cycles/month. In the context of smoothing a monthly input, the frequency domain $\Omega = (0 < \omega < 0.5)$ can be portioned in two main intervals: (1) $\Omega = (0 \leq \omega \leq 0.06)$ associated with cycles of 16 months or longer attributed to the signal (trend/cycle) of the series and (2) $\bar{\Omega} = (0.06 < \omega < 0.5)$ (the signal of higher frequencies up to the Nyquist frequency) corresponding to short cyclical fluctuations and the noise. It is in particular for the signal in this range that the full strength of the auto-regressive modelling offered, for instance, by X-13-ARIMA can be employed to further characterise the time series for in-depth analysis. Figure 2.1 shows an example of what can be achieved in terms of a filter designed to separate out signal in a band of frequencies between $\sim 1/12$ cycle/month and $\sim 5/12$ cycles/month. Evidently this is not completely unique, in the sense that slight variations in this frequency response that are visually almost un-noticable, would have some effects on the weights w_k . The weights w_k that correspond

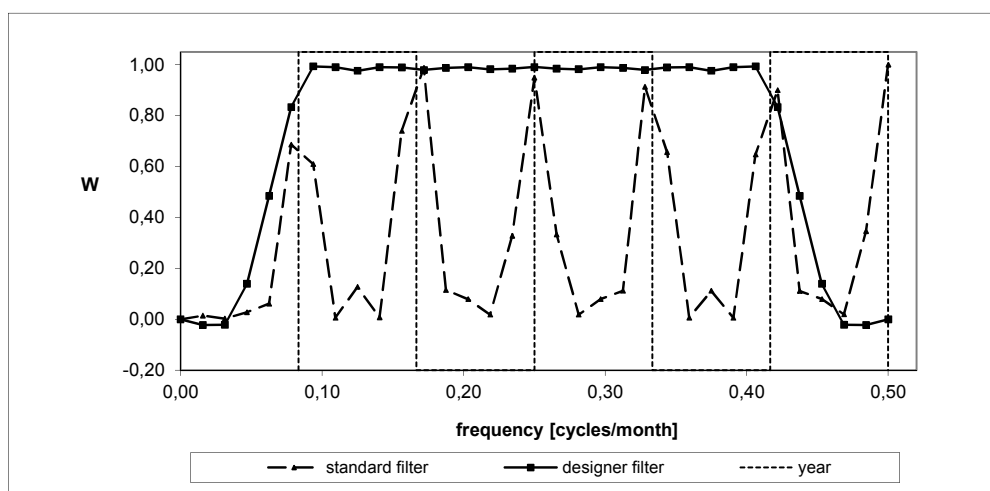


Figure 2.1 The band-pass filter designed for seasonal adjustment. Dashed line indicates the standard filtering. Small dashed line indicates integer multiples of 1 cycle/year.

to this filter of Figure 2.1 are shown to 7 digits precision in Table 2.1. The weights w_k for all the odd values of k are identically equal to 0, as are the weights for $k > 18$. This set of weights satisfies the various constraints mentioned in this section. The additional property that the weights are zero for all odd values of k produces an additional advantage: the high frequency

section of the spectrum can be determined in a very simple second step, which is evidently statistically independent. If y is the time series from which the seasonal component has been removed with the above filter, then the high frequency component h is obtained by taking:

$$h_i = (2\tilde{y}_i - \tilde{y}_i - \tilde{y}_{i+1})/4 \quad (13)$$

Whereas the *trend+cycle* time series c (without high frequency contributions) is obtained from:

$$c_i = (2\tilde{y}_i + \tilde{y}_i + \tilde{y}_{i+1})/4 \quad (14)$$

With the weights of Table 2.1 it is clear that after 18 months any seasonal adjusted time series y will not change, when using this filter alone.

Table 2.1 Filter weights for a band-pass filter designed to extract seasonal behaviour from a time series.

k	$w_k = w_{-k}$
0	0,7358026
2	-0,2219532
4	-0,1504270
6	-0,0659661
8	0
10	0,0309203
12	0,0302373
14	0,0143577
16	0
18	-0,0050703

Given that applying this filter to any given time series is as straightforward as the more standard filtering steps (such as applying Henderson filter for smoothing a time series) there is no fundamental difficulty in incorporating this weight system in the standard packages that are currently widely available. By doing this, for instance, with the X-13-ARIMA package, all the other tools that this package offers are still available. In particular, the high-frequency “noisy” term h would be the most interesting target for further ARIMA modelling. A possible alternative is to

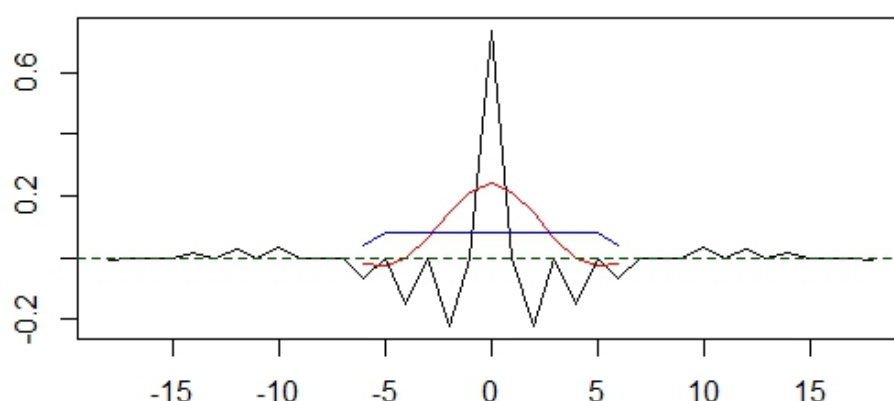


Figure 2.2 Time-domain representation of different filters. Black line is the designed filter, ranging in a [-18:18] terms width; red line indicates the Henderson 13-terms filter and blue line indicates the 13-terms seasonal moving average filter.

pre-process time series with this filter, and this is exactly our proposed procedure. There is no loss of information in this pre-processing step since the seasonal factor and the *trend+cycle+noise* series are both available and their sum is identical to the original time series. The time series filtered with the proposed weight system can then be provided as input for the standard packages (in this case, JDemetra+), in order to be able to use the extensive toolset of those packages for a more detailed analyses of the character and causality of the seasonal behaviour. Such more in-depth analyses could be “off-line” and would not interfere with the normal publishing schedules of the original and seasonal adjusted time series. Figure 2.2 plots the two mentioned set of filters represented in the time domain, comparing them with the function of the designed one.

2.3 Periodogram

Time series are generally presented in the time domain. Another interesting representation of the same series is in the frequency domain, considering that any stationary time series can be expressed as a combination of sinusoidal functions, thanks to the Fourier transform. It represents the spectrum, and its representation is called the periodogram. It is possible to obtain it as follows:

$$y_t = \sum_{i=1}^{\frac{n}{2}} [\beta_{1_i} \cos(2\pi\omega_i t) + \beta_{2_i} \sin(2\pi\omega_i t)] \quad (15)$$

The β 's work as regression parameters, obtained from: $\beta_{1_i} = A_i \cos(\phi_i)$ and $\beta_{2_i} = -A_i \sin(\phi_i)$, where ϕ is called the phase. It determines the starting point (in degrees) for the cosine wave. A is the amplitude.

The spectral analysis investigates the frequency domain representation of the series to determine how important cycles of different frequencies determine the behaviour of y_t . Therefore, the periodogram is useful to identify the dominant cycles of a time series. Peaks in a periodogram indicate frequencies of cyclical movements with a certain periodicity. The periodicity P of a phenomenon is related to its frequency ω : $P = 2\pi/\omega$. It means that for a monthly series, seasonal frequencies are $\pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6$ and π , and they correspond to 1, 2, 3, 4, 5 and 6 cycles per year (for quarterly sampled series, the two seasonal frequencies are $\pi/2$; one cycle per year, and π ; two cycles per year).

The trading day frequency is 0.348, since a daily component which repeats every seven days goes through 4.348 cycles in a month of average length 30.4375 days. It means, then, that 0.348 is the trading days frequency for monthly sampled time series, considering the average length of a year (365.25 days). Actually, the original trading days frequency (4.348) is higher than the Nyquist frequency. If any signal has a frequency higher than the ω_{Nyq} , it is an aliasing signal, i.e. it is indistinguishable from a lower-frequency signal. When this happens, the signal “appears” with much lower ω . Consequently, $\omega = 4.348$ appears as 0.348.

A time series with a strong seasonal component should have peaks at the seasonal frequencies, whereas a seasonally adjusted time series should have no peaks at those frequencies. Moreover, the presence of the trend component is always evidenced by a peak at frequency zero and so, when a periodogram has high values at low frequencies, i.e. at $(0 \leq \omega \leq 0.06)$, it means that the long-term component (the trend) is dominating the series. On the contrary, if high values of the periodogram concentrate more at high frequencies, the time series is rather trendless and has a significant irregular component.

JDemetra+ contains different spectral diagnostics. First, it is possible to obtain a spectral plot of the raw time series, obtained (by default) after the performance of a first difference operation on

the time series, to get the data stationary. Indeed, non-stationarity of a series could lead to a misinterpretation of the periodogram. One other issue is related to the presence of outliers, since they can influence the outcome of the periodogram as well. This is why a pre-transformation of data is advisable, as done by both TRAMO/SEATS and X-13-ARIMA. As is visible from Figure 2.3, in JDemetra+ the seasonal frequencies are marked as grey vertical lines, while the bordeaux lines represents the trading day frequency. In addition to the 'raw' periodogram, once the seasonal

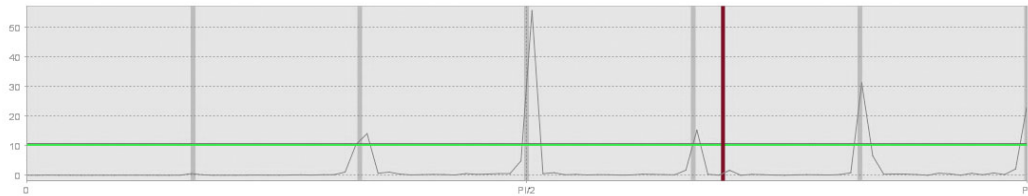


Figure 2.3 JDemetra+ representation of a periodogram.

adjustment procedure has been carried out, JDemetra+ offers some spectral diagnostics. Especially, it is possible to obtain periodograms of the seasonally adjusted series, residuals and irregular component. Further details will be given in the next section.

3 Seasonal adjustment procedure

The time series used in this study represents the air passenger movements registered at the Schipol airport in Amsterdam, the Netherlands. The series covers a length of 196 monthly registrations, spanning the period from January 1999 to April 2015. As illustrated in Figure 3.1, the data present a strong seasonal pattern, which is slightly varying over the entire series.

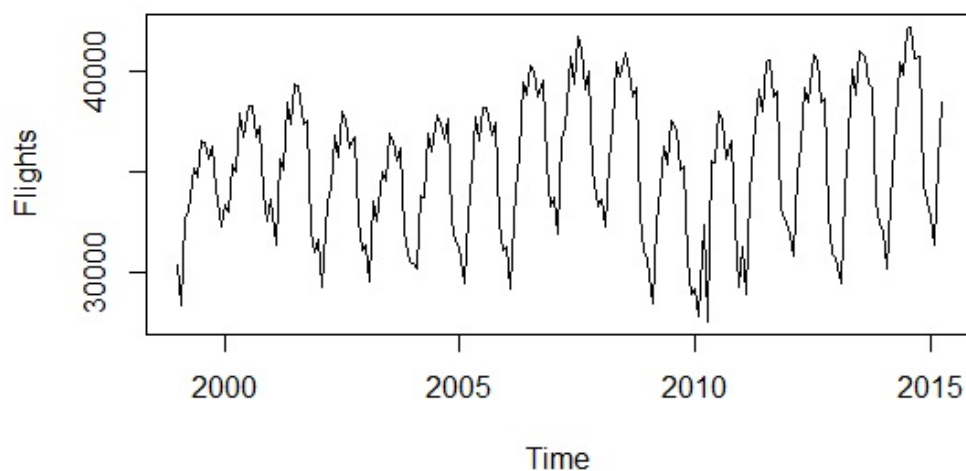


Figure 3.1 Amsterdam Schipol airport flights from January 1999 to April 2015.

3.1 JDemetra+

The first step is to carry out a seasonal adjustment with JDemetra+, i.e. with TRAMO/SEATS and X-13-ARIMA procedures, without any pre-filtering treatment of the series. The software offers the possibility, through its *workspace* window, to choose between different pre-defined *specifications* implemented in the two methods. Those *specifications* are sets of parameters and values assigned to them that contain all information necessary for seasonal adjustment and for modelling the time series during the pre-decomposition step. Table 3.1 reports all the possible specifications (the first seven are about TRAMO/SEATS, the last six about X-13-ARIMA), but it is important to highlight two main points. First, the user has the option to set manually those parameters, depending on the nature of the analysis. Second, this paper will focus on the default specifications settled by JDemetra+ (*RSAfull* for TRAMO/SEATS and *RSA4c* for X-13-ARIMA), but with a fix model parameters choice, the so called ARIMA “Airline Model” $(0,1,1)(0,1,1)$ (Box and Jenkins, 1976), one of the most commonly used seasonal ARIMA model, in order to meet the parsimony principle. The *Transformation* column indicates the performance of a test to choose between an additive decomposition (no transformation) and multiplicative decomposition (logarithmic transformation) of the time series. Concerning the ARIMA model selection, the acronym AMI stands for Automatic Model Identification. A detailed discussion of different parameters and specifications can be found in (Grudkowska S., 2015). The software performs an

Specification	Transformation	Pre-adjustment for leap-year	Working days	Trading days	Easter effect	Outliers	ARIMA model
RSA0	no	no	no	no	no	no	$(0,1,1)(0,1,1)$
RSA1	test	no	no	no	no	test	$(0,1,1)(0,1,1)$
RSA2	test	no	test	no	test	test	$(0,1,1)(0,1,1)$
RSA3	test	no	no	no	no	test	AMI
RSA4	test	no	test	no	test	test	AMI
RSA5	test	no	no	test	test	test	AMI
RSAfull	test	no	test		test	test	AMI
X11	no	no	no	no		no	$(0,1,1)(0,1,1)$
RSA1	test	no	no	no		test	$(0,1,1)(0,1,1)$
RSA2c	test	test	test	no	test	test	$(0,1,1)(0,1,1)$
RSA3	test	no	no	no		test	AMI
RSA4c	test	test	test	no	test	test	AMI
RSA5	test	test	no	test	test	test	AMI

Table 3.1 TRAMO/SEATS and X-13-ARIMA pre-defined specifications.

automatic identification of the calendar effects of a time series, and includes them as regressors in the RegARIMA model fitted in the *modelling* part. This paper will not offer a thorough discussion about these regression parameters, i.e. outliers, Easter effect, trading/working days effect, leap-year effect; for it will mainly focus about the Fourier-domain representation of the historical data. A useful discussion of the trading days effect identification and its spectral diagnostic is offered by (Findley D.F., Soukup R.J., 1999). In this case study, this determinist effects identification will not be modified from the automatic one performed by the software.

3.2 JDemetra+ spectral analysis

A helpful tool implemented in JDemetra+ refers to the spectral analysis of time series. The possibilities for a spectral analysis offered by the software are Periodogram, Auto-regressive Spectrum and Tukey Spectrum. This paper will focus only on the periodogram. Figure 3.2 presents the periodogram of the analysed historical data. As mentioned in Section 2, the JDemetra+ periodogram considers a first difference operation on data. As in most of the

JDemetra+ tools, users are free to change this setting, by increasing the differencing order or deleting it. JDemetra+ points out the significance of a peak by a green horizontal line, which denotes the 0.05 significance level. The unadjusted series shows a peak at the first seasonal

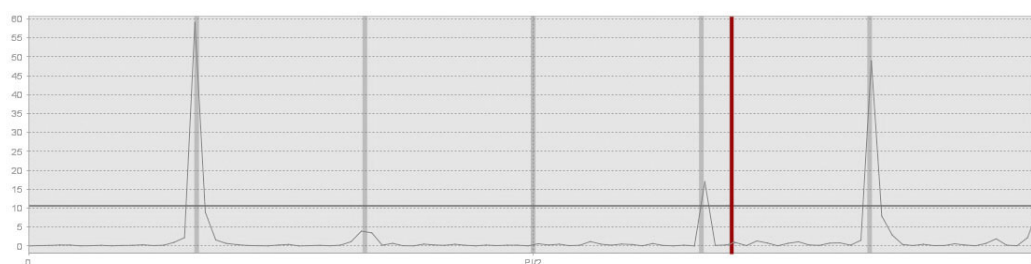


Figure 3.2 Periodogram of unadjusted time series.

frequency, which means the presence of a strong seasonal movement of one cycle per year. Another significant peak coincides with the fifth seasonal frequency, related to a movement of five cycles per year. This periodogram gives results which could be visible also from the simple plot of the series (the presence of seasonality), but from this frequency-domain representation it is more straightforward to check the cycles of the dominant seasonal movements. When starting a seasonal adjustment with JDemetra+, both the methods imply two steps; a *modelling* part and a *decomposition* part. As briefly discussed in the introduction, the *modelling* part consists of identifying the regression parameters that have a deterministic effect on the data, and then of fitting an ARIMA model to the data, to give a suitable model for an acceptable decomposition. Since this paper focuses on the designed filter to decompose a time series in its components, it does not offer detailed discussion of the *modelling* results, but rather on the spectral diagnostics of the decomposition output and the filtering procedures. For an exhaustive discussion about the *modelling* part see (Maravall A., 2008b). Indeed, JDemetra+ offers several diagnostics, using a user-friendly graphic for the results of the various tests. Together with every *p.value* associated to each test, an evaluation of the result is given in words (good, uncertain, severe) coloured, respectively, in green, yellow and red. This helps for interpretation of the results.

3.3 TRAMO/SEATS and X-13-ARIMA spectral diagnostics

Once the seasonal adjustment has been performed, JDemetra+ evaluates the quality of the decomposition through a wide range of diagnostics. Within these, a *spectral diagnostic* consists in three frequency-domain charts, one for the model residuals, one for the irregular component, and one for the seasonally adjusted series. Actually, the software offers these diagnostics both as periodogram and as auto-regressive spectrum, but only the periodograms are considered here. Figure 3.3 plots the periodogram of the residuals. A peak in this periodogram at seasonal or trading day frequencies indicates the need for fitting a better model. Particularly, peaks at seasonal frequencies suggest a different filter choice for the decomposition of the series, while peaks at the trading day frequency suggest the inappropriate identification of the deterministic effects used as regression variables in the model. In the periodogram of residuals obtained from the TRAMO/SEATS method no peak exceeds the significance level, even if some movement is present through different frequencies; particularly, a peak is visible at the first seasonal frequency. A peak in the spectrum of the seasonally adjusted series or in the spectrum of irregular component reveals the inadequacy of the seasonal adjustment filters for the time interval used for the spectrum estimation. In this case a different model specification or a different data span length should be considered.

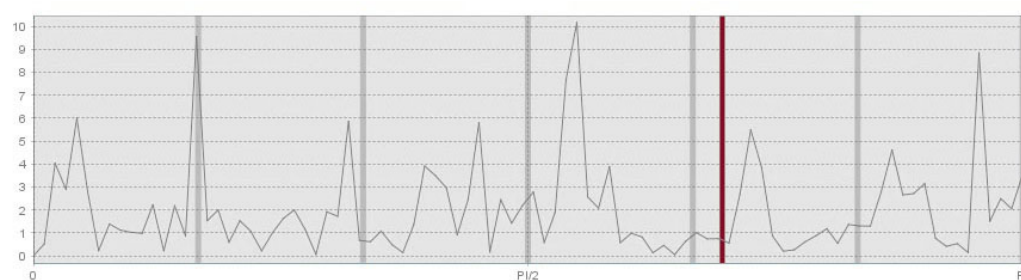


Figure 3.3 TRAMO/SEATS residuals periodogram.

For computational reasons, in JDemetra+, none of the periodograms for seasonally adjusted series and for the irregular component displays the horizontal green line that highlights peaks above the 0.05 significance level. In the adjusted series (figure 3.4, lower chart), the periodogram does not show any peak at seasonal frequencies, reflecting a good decomposition of the original series performed by SEATS and a satisfactory removal of the seasonal effect. As desirable, low frequencies have low amplitudes, while higher values are registered at higher frequencies, in between the seasonal harmonics. Indeed, the noise component is not totally removed, as can be seen in the upper chart of Figure 3.4, related to the irregular periodogram, which shows relevant power between seasonal frequencies. In the next section, about the

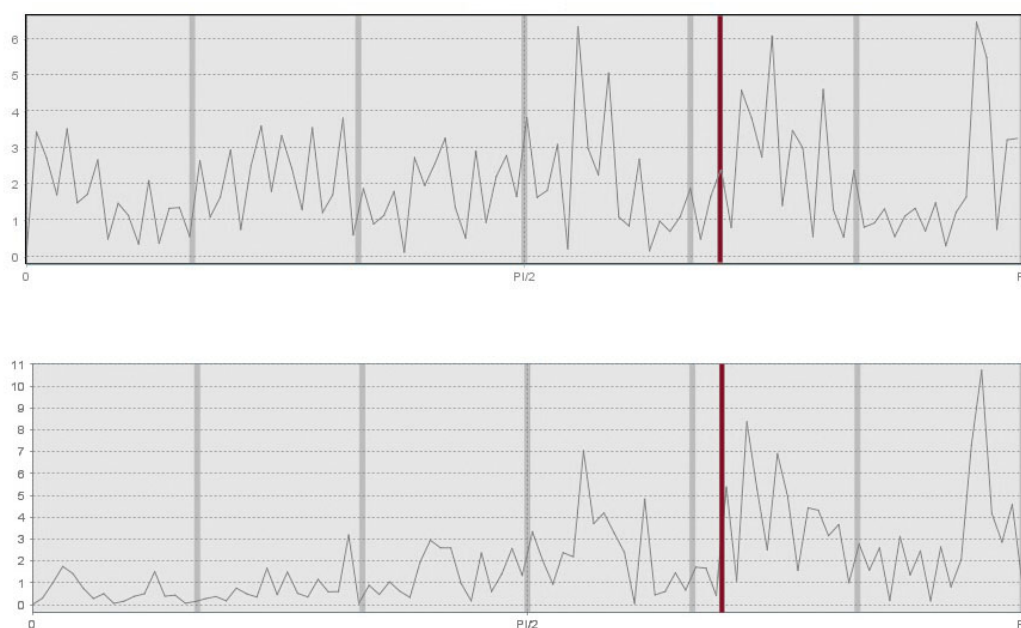


Figure 3.4 TRAMO/SEATS irregular and seasonally adjusted series periodogram.

seasonal adjustment of the filtered series, these same plots will show a much smoother behaviour. In the following charts are reported the same three periodograms, but the ones obtained from the X-13-ARIMA procedure, which makes the use of 3x3 seasonal filter and a 13-terms Henderson moving average as trend filter. The length of these filters depends on the time series, and the selection is done automatically by the X-11 algorithm (Dagum E.B., Chhab N., Chiu K., 1996). For the comprehension of the length selection criterion, see also (Grudkowska S., 2015), while a brief discussion about how these filters work will be provided in the appendix. Periodograms in Figure 3.5 show a similar behaviour to the ones obtained from TRAMO/SEATS, and their interpretation is the same.

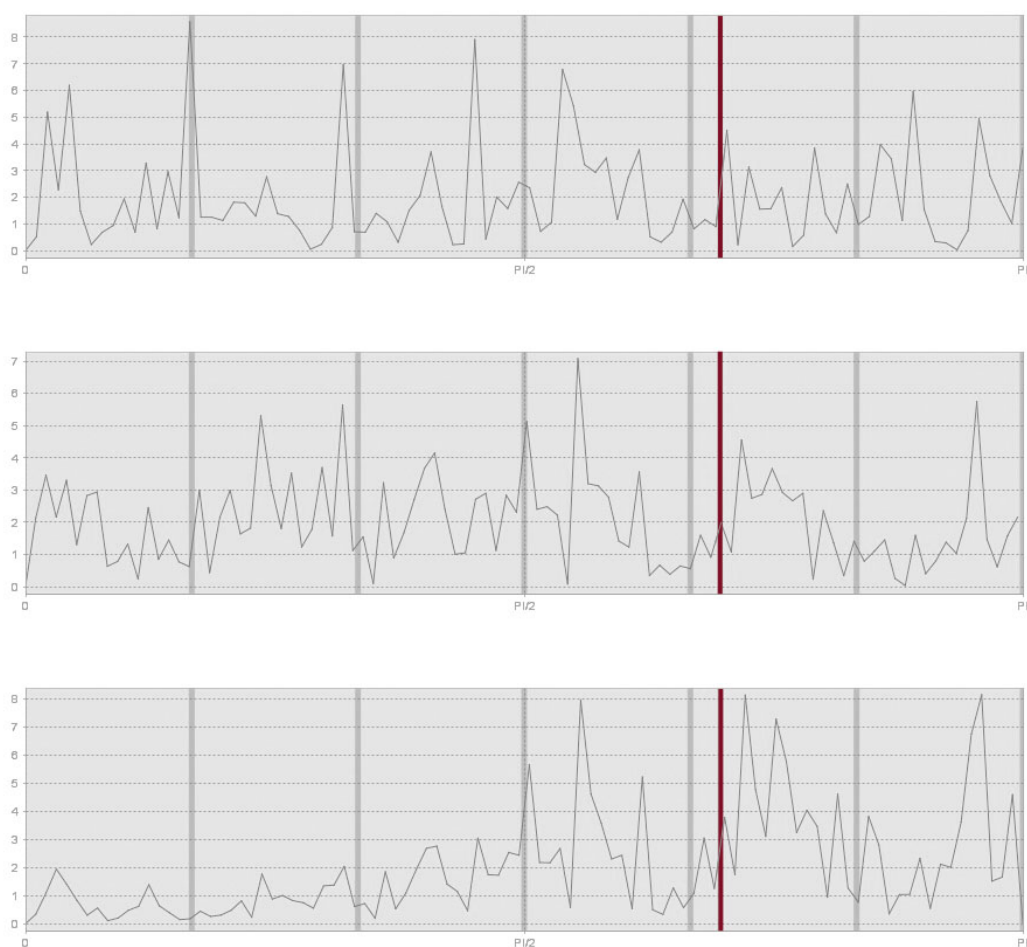


Figure 3.5 X-13-ARIMA residuals, irregular and seasonally adjusted series periodogram.

3.4 Squared gain of components filter.

TRAMO/SEATS estimates the various components using the so-called Wiener-Kolmogorov filters. These filters are symmetric. It means that observations from the past and from the future have to exist, so there is the need for forecast and backcasts for the both ends of the series. This operation is performed through the ARIMA model estimated by TRAMO. Further discussion about these filters is provided in the Appendix.

An interesting tool related to the quality of the decomposition is the square gain of the components filter. This indicates which frequency components are suppressed or amplified by the filter, i.e. it controls which movement of particular amplitude at a frequency ω is delivered to the output series (derivation of the squared gain function is briefly discussed in the appendix). If in a frequency band $[\omega_1; \omega_2]$ the squared gain has value zero, then the output series is free of movements in that range; if is equal to one, all the movements from that range is passed on to the component estimator. It means that the squared gain of the seasonal component estimator filter is expected to capture only the seasonal frequencies (taking unitary values at these frequencies), while for the seasonally adjusted series, this function should have value equal to zero at those same frequencies, in order to eliminate the seasonal effects. In general, the squared gain shape depends on the model for the time series. A thorough explanation of the derivation of the squared gain function and its interpretation, as the values of the parameters of the ARIMA model and the length T of the series change, is discussed in (Findley D.F., Martin

D.E.K., 2006). Figure 3.6 represents the squared gain of the components filter obtained by

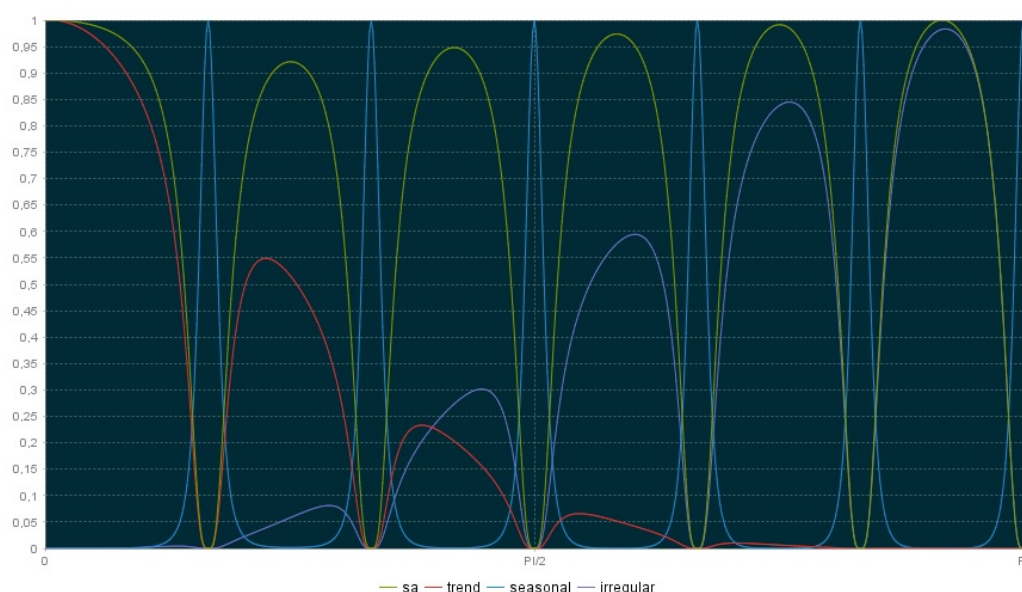


Figure 3.6 Squared gain of TRAMO/SEATS components filter

TRAMO/SEATS from the dataset analysed in here. As the value for seasonal component estimator is equal to one at seasonal frequencies, the decomposition procedure seems to act quite well, with relatively narrow troughs. On the other hand, the shape of the squared gain of the seasonally adjusted component filter seems to miss some signal between seasonal frequencies, and it does not reach the exact value of one. It will be interesting, in the next section, to compare this squared gain chart to the one obtained from the same time series, but pre-filtered with the proposed filter. As X-13-ARIMA makes use of the algorithm of X-11, which implies the use of standard filters (seasonal moving average and Henderson filters, where only the filter length may vary), JDemetra+ does not offer a squared gain representation of these filters, but only for the TRAMO/SEATS ones, since they are defined differently for each analysed time series.

4 Filtering a time series with the designed filter

4.1 First iteration of the designed filter on a time series

The filtering of the time series using the filter proposed in Section 2 has been performed with the statistical software R. The decomposition of the time series through the filtering operation is represented in Figure 4.1. The upper-left panel of Figure 4.1 represents the seasonal component, obtained with a single iteration of the designed filter through equation (7). Then, this component has been subtracted from the original series, in order to obtain the *trend+cycle+noise* series, represented in the upper-right panel (blue line) in a comparison with the unadjusted data (black line). It is worth to highlight again that the sum of this *trend+cycle+noise* and the seasonal factor is identical to the original time series. According to equation 14, it is then straightforward to identify the high-frequency (noise) component and then to subtract it from this

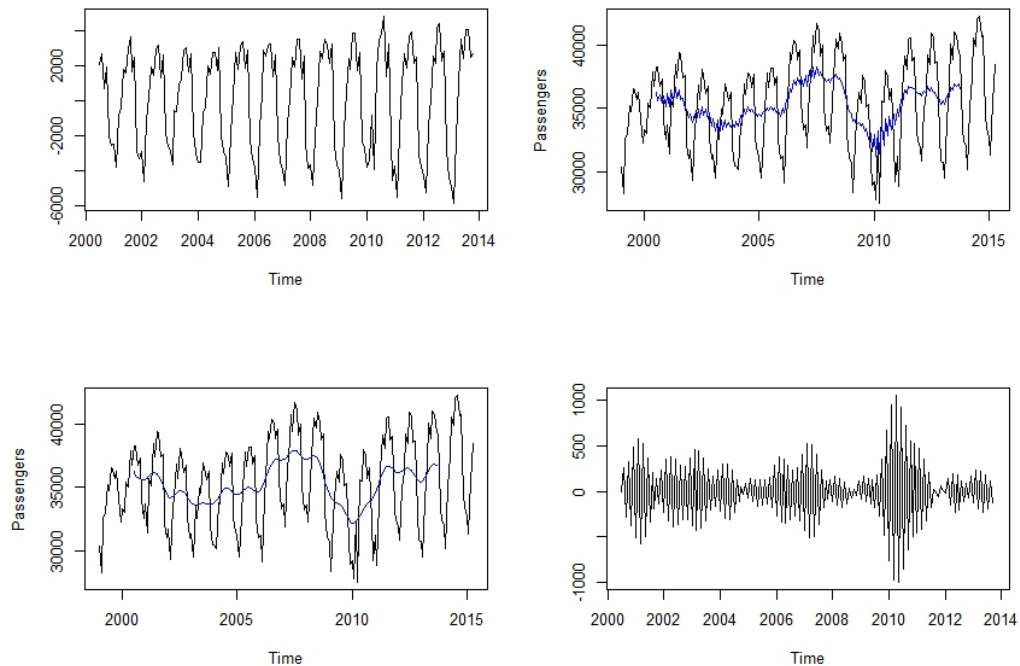


Figure 4.1 Decomposition of the time series with the proposed filter.

trend+cycle+noise series. The pure smooth *trend+cycle* component (the one in blue as well) is compared again to the original data (black line) in the bottom-left panel. At last, bottom-right chart reveals the high frequencies component, i.e. the “noise” of the series, which shows a strong increase in amplitude in the year 2010. Actually, an explanation of this increase could be the two additive outliers detected by both TRAMO/SEATS and X-13-ARIMA in this year, in the months of April and of December.

The outcome of the seasonally adjusted data is not totally satisfactory. While clearly a great amount of the seasonal pattern is removed, the visual impression is that the filtering has not been completely successful. One of the reasons for this is the strong amplitude of the seasonal component. One could deal with this in a number of different ways. One possibility is simply to repeat the filtering process, but since the aim of this study is the analysis of the output of the seasonal adjustment procedures with the filtered series as input, the second filtering operation will be not performed in here. Figure 4.2 plots the JDemetra+ periodogram of the time series filtered, through R, with the proposed filter, before the performance of the further seasonal adjustment. This time, as opposed to what was done in Section 3, there is no need for differencing the time series, since it is already stationary. Therefore, the first difference operation has been removed

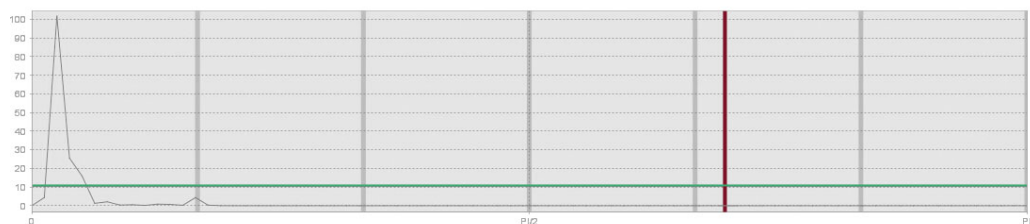


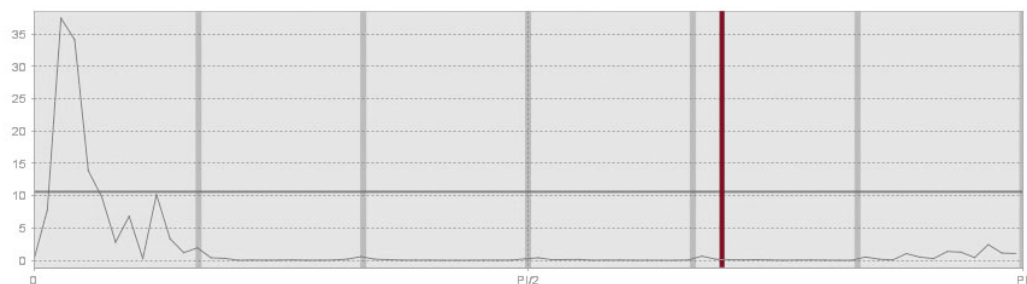
Figure 4.2 Filtered time series periodogram.

As a proof of the imperfect seasonality removal, a peak is visible at the first seasonal frequency

of 1/12 cycle/month. This imperfect removal after one pass of the filter is due to the fact that the frequency of 1/12 cycle/month is near the edge of the band which the seasonal filter removes. The removal is therefore slightly less than 100%, which will leave a noticeable residual for a signal with a very strong annual variation. One could try to deal with this by making the band to be removed slightly broader, but this would imply that there would also be some undesirable suppression of genuine longer term cycle behavior. Another option might be to try to produce a sharper transition in the band-pass filter from the region in the spectrum where it passes all signal to the region in the spectrum where all signal is blocked. While this is possible, the effect it has is that in the time domain the linear coefficients of the filter function are non-zero over a larger range. In practice there is always a trade-off between the level to which seasonal behaviour can be suppressed in the time series for *trend+cycle+noise*, and the range over which the filter function has non-zero values. Apart from this limitation, the periodogram has values equal to zero for all the frequencies after the first seasonal one. It means that the removal of the seasonal component through the designed filter works properly for all the other harmonics. To confirm the efficacy of the seasonal removal, it is possible to compare this periodogram to the one in Figure 3.4 and 3.5, of the seasonally adjusted series obtained through TRAMO/SEATS and X-13-ARIMA.

4.2 Filtered series as input for JDemetra+ analysis

The filtered time series (the *trend+cycle* series represented in bottom-left chart of Figure 4.1) has been used as input for a JDemetra+ analysis. On some occasions, TRAMO provides a not decomposable model. It happens when components have a negative spectrum for some frequencies. It is then said that the model presents a “non-admissible” decomposition. When this happens, SEATS automatically modifies the model parameters, searching for a decomposable model that is not far from the one identified by TRAMO. The search always converges. For a more extensive discussion, see (Maravall A., 2008a). This is the case of the TRAMO/SEATS analysis performed on the filtered series, which finds the TRAMO model not decomposable. As mentioned in Section 3, the considered models is the (0,1,1)(0,1,1) ARIMA model. Before comparing the spectral analysis of the outputs, it is worth to compare the seasonal factors obtained from the JDemetra+ decomposition performed on the filtered data and on the non-filtered data. Relevant is the difference in the amplitude of the seasonal factor. The lower chart of Figure 4.3, of the not filtered data, shows an amplitude ranging between -6000 and 4000. The seasonal factor component of the filtered data (upper chart) shows a trend either smoother (at the peaks) and lower, ranging between -400 and 400. In the following charts are represented the spectral diagnostics of the output of the TRAMO/SEATS method, obtained when using the filtered series as input for the seasonal adjustment.



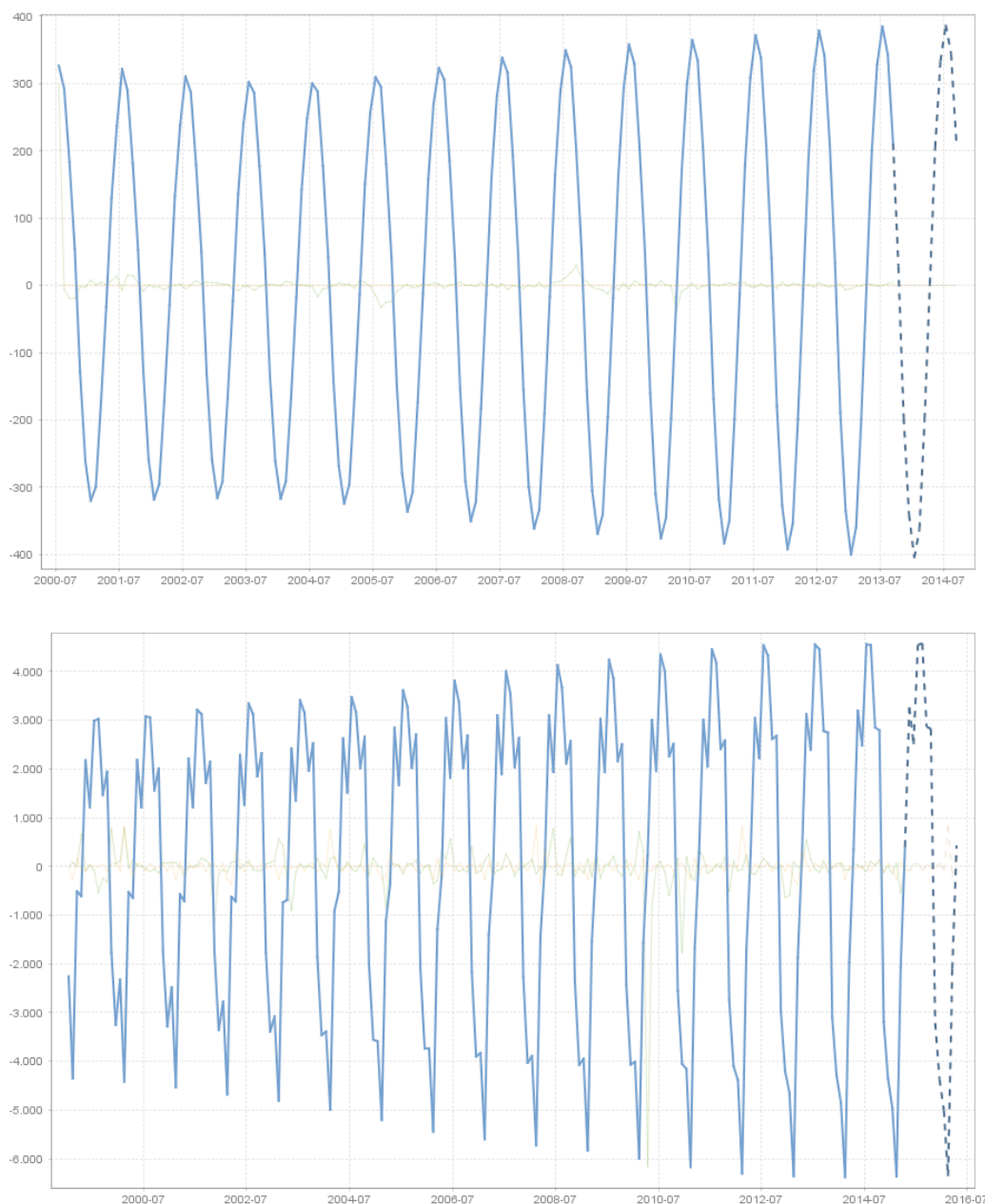
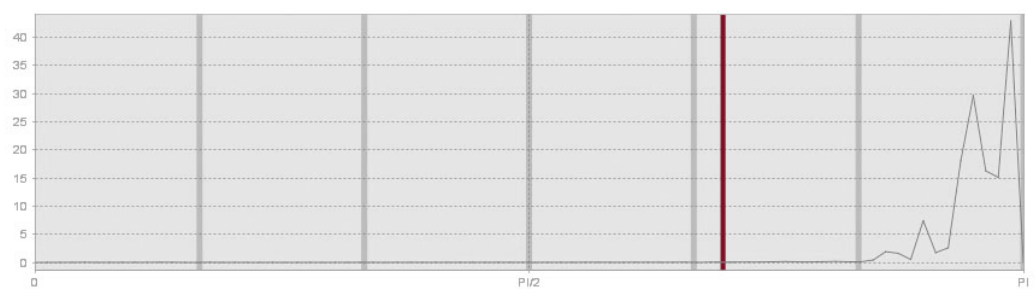


Figure 4.3 Comparison of seasonal factors obtained through X-13-ARIMA decomposition (TRAMO/SEATS acts similar). First is the filtered data seasonal component, second is the non-filtered data seasonal component.



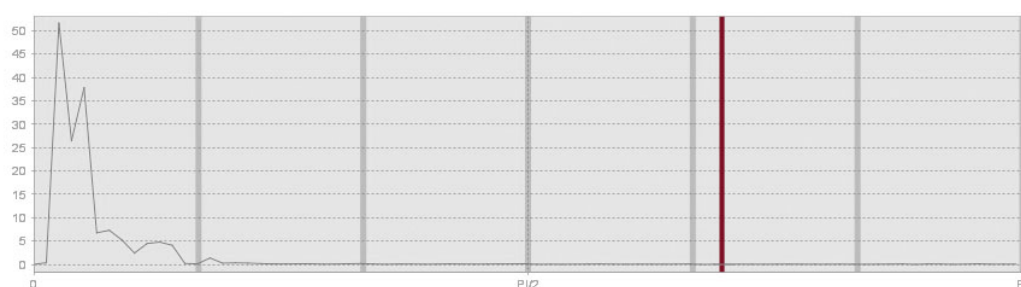


Figure 4.4 Residuals, Irregular and seasonal adjusted periodogram obtained by TRAMO/SEATS.

Most relevant result about these periodograms is the total absence of power at any seasonal frequency. This is a proof of the efficacy of the designed filter, particularly when the filtered series is used as input for further analysis. The comparison between Figure 4.4 and the periodograms analyzed in Section 3 confirms this. Periodograms obtained from X-13-ARIMA show a similar behaviour to the TRAMO/SEATS ones, and therefore they will not be reported in here. JDemetra+ offers many other diagnostics. Some of them are related to the presence of seasonality. As a proof of the removal of the seasonal component, Table 4.1 represents the results of these tests (for both the procedures), whose check for the presence of seasonality in the seasonally adjusted series. An explanation of these tests is offered by (Grudkowska S., 2015). As desirable, they testify the efficient removal of the seasonal factor from the time series. When

Test	Seasonality	Test	Seasonality
1. Auto-correlations at seasonal lags	NO	1. Auto-correlations at seasonal lags	NO
2. Friedman (non parametric)	NO	2. Friedman (non parametric)	NO
3. Kruskal-Wallis (non parametric)	NO	3. Kruskal-Wallis (non parametric)	NO
4. Spectral peaks	?	4. Spectral peaks	NO
5. Periodogram	NO	5. Periodogram	NO
5bis. Max Periodogram	NO	5bis. Max Periodogram	NO
6. Seasonal dummies	NO	6. Seasonal dummies	YES

Table 4.1 JDemetra+ output for tests about the presence of seasonality on seasonally adjusted series; TRAMO/SEATS on the left and X-13-ARIMA on the right.

fitting a model to a time series, one of the most reliable indicator of the goodness of the fit is the Akaike Information Criterion (Akaike, 1974). When choosing between two different models, it is preferable to use the one with the lowest AIC value. This value in the two X-13-ARIMA procedures (the first one for the unadjusted data, the second one for the filtered data) is significantly different: it is respectively 2842.46 for the unadjusted time series and 1761.82 for the filtered time series.

Figure 4.5 shows the squared gain of the components filters used by TRAMO/SEATS for the filtered time series decomposition. It is perfectly visible that the square gain of the seasonally adjusted filter catches much more signal if compared to the one of the unadjusted data, visible in Figure 3.6, for it has value equal to one between the seasonal frequencies. It means that the seasonal component of the filtered data is much more deterministic than the one of the original data. To confirm this, the bright blue line, related to the seasonal factor filter, has value equal to zero for all the frequencies out of the seasonal ones. Furthermore, the squared gain of the trend filter has higher values for all the frequencies compared to the squared gain of the trend filter of Figure 3.6. It means that all information about long term behavior is delivered to the output series.

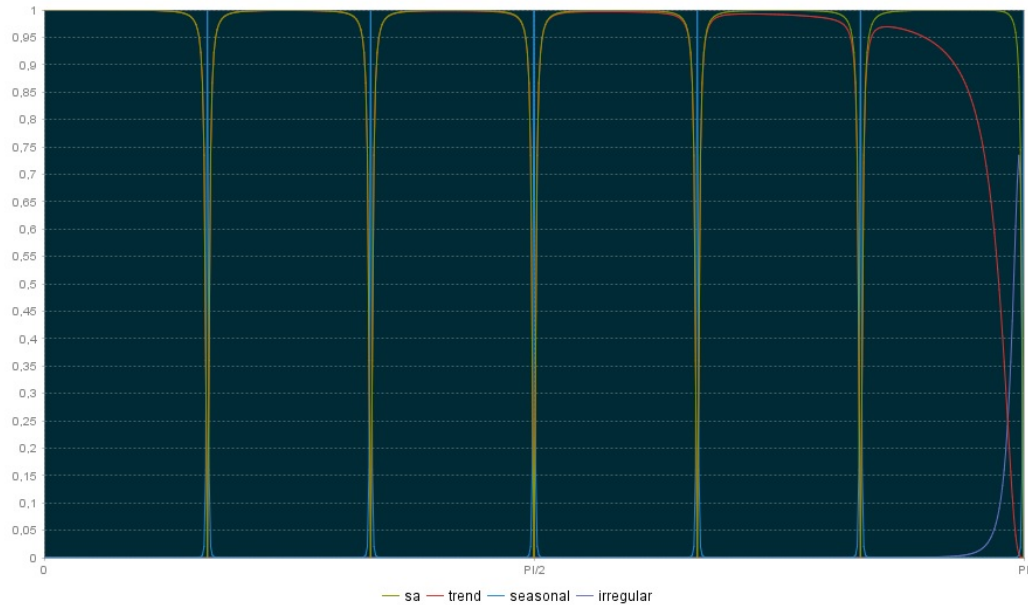


Figure 4.5 Gain filter of the components of TRAMO/SEATS procedure applied to filtered time series.

5 Conclusions

It is possible to design linear filters for seasonal adjustment of time series that are more effective for real-time applications in official statistics than the filters currently in widespread use. Most packages already provide several options for filter coefficients, which means that it is straightforward to add the filter presented in this paper as another alternative.

The advantage of the filter presented here is that the coefficients are non-zero over a range that is rather more restricted than the one obtained when recursively applying the existing naive options. This means that the resulting time series with seasonal adjustments can be finalised faster, and extrapolation towards the end points of time series is better constrained since it needs to take place over a much more limited range.

An additional benefit of filtering all signals in a frequency band, is that this adjusts to a large degree for slow modulation of the amplitude of seasonal effects, as well as compensating irregularities in monthly sampling which may occur because in practice the sampling is done eg. on the last friday or the last working day of the month even though it is treated as if it were perfectly regular.

The linearity of the filter together with the non-recursive formulation of it implies that partial time series can be adjusted for seasonal influences separately and then totalised which will by construction yield identical results to the seasonal adjustment of the summation of the partial time series.

A JDemetra+ analysis of the output of this filtering operation confirms the aimed achievements.

Appendix

Standard filtering

The standard initial filtering steps normally taken within the X-12-ARIMA package are threefold. In the first step a provisional trend is identified by taking a moving average with the uniform weights and end-point correction, i.e. if the trend and cycle are indicated by c_i then:

$$c_i = \frac{1}{24}y_{i-6} + \frac{1}{12}y_{i-5} + \dots + \frac{1}{12}y_{i+5} + \frac{1}{24}y_{i+6}$$

In some literature this is referred to as $M_{2 \times 12}$ filter. This provisional trend is then subtracted from the time series to produce a provisional *seasonal+irregular* component. At this stage the time series can be investigated for outliers which can be downweighted in a variety of ways. The next step is for each calendar month to calculate a moving average using the same month in several successive years. Often most weight is given to that month in the “central” year and symmetrically decreasing the weights for earlier or later years. Other options are available however.

An updated *trend+cycle+noise* series c can now be computed, by subtracting the seasonal component that has just been determined from the original time series. This can be improved upon by once again applying a moving average filter, but rather than a uniform moving average, a Henderson filter is applied, which tapers more smoothly over the time series. In this way an updated *trend+cycle+noise* is obtained which is the starting point also for obtaining an updated seasonal component. In principle one can iterate this process further.

The combination of these individual filtering steps can also be expressed in terms of a single set of filtering weights since each individual step is simply a linear combination of the data with known weights. The combined steps described above tend to lead to a set of weights which is non-zero over a considerable range of months: $[-84, 84]$ although the magnitude of these weights is generally only appreciable for the range $[-40, 40]$.

The implication of this is that for at least 40 months after any given reporting month the seasonal adjusted time series cannot be considered final, or even for 84 months if the full formal range of non-zero weights is taken into consideration. Further autoregressive modelling, if applied, will exacerbate this problem.

Wiener-Kolmogorov filter

The aim is to extract an estimate of a signal sequence $s(t)$ from an observable time series $y(t)$:

$$y(t) = s(t) + \eta(t)$$

where $\eta(t)$ is called noise. TRAMO/SEATS decomposition procedure works by means of the following algorithm.

An ARIMA model is fitted for the signal:

$$\phi_s(B)s_t = \theta_s(B)a_{st} \quad \text{where} \quad a_{st} \sim w.n.(0, V_s)$$

where the model for the time series is:

$$\phi(B)y_t = \theta(B)a_t \quad \text{where} \quad a_{st} \sim w.n.(0, V_a)$$

Notice: $\phi(B) = \phi_s(B)\phi_\eta(B)$, where B is the backshift operator ($B^j y_t = y_{t+j}$), and the a_t are the white-noise innovations.

The two polynomials $\phi(B)$ and $\theta(B)$ are, respectively, the stationary autoregressive polynomials in B and the invertible MA polynomials in B .

Write:

$$s_t = \Psi_s(B)a_{st}; \quad \Psi_s(B) = \frac{\theta_s(B)}{\phi_s(B)};$$

$$y_t = \Psi(B)a_t; \quad \Psi(B) = \frac{\theta(B)}{\phi(B)};$$

The SEATS estimation of the signal s at time t then is:

$$\hat{s}_t = \left(\frac{V_s \Psi_s(B) \Psi_s(F)}{V_a \Psi(B) \Psi(F)} y_t \right) = v(B, F) y_t = \left(v_0 + \sum_{j=0}^{\infty} v_j (B^j + F^j) \right) y_t$$

Where F is the forward operator; $F = B^{-1}$; $F^j y_t = y_{t+j}$, and $v(B, F)$ is the so-called Wiener-Kolmogorov filter (WK).

Note: if time series is stationary, WK filter is equal to:

$$v(B, F) = \frac{ACGF(s_t)}{ACGF(y_t)}$$

Wiener-Kolmogorov filters theory is based on the MMSE (Minimum Mean Square Error) idea. The filters derive from the observed model, which in turn depends on the time series. So, those filters differ depending on the observed series. WK filters are symmetric and bi-infinite; it means that for the ends of the series there is the need of forecasts and backcasts in order to get the estimations of those time points as well. This is performed by the TRAMO phase of seasonal adjustment, which provides an extended ARIMA model fitted to the data. Then, SEATS applies the WK filters. In this way, at the end of the time series the estimator is preliminary and is subject to revisions every time a new observation enters the time series.

Since the filters are centered, they are also convergent. It enables to approximate an infinite number of realization by a finite number of terms. With this procedure, highest weights are applied for central observations, even if the weighting pattern depends on the component to estimate. For the seasonal component, for example, highest weights are applied to the current value and the past and future values from the same period.

Square Gain of a filter

In the context of filtering a series, it is possible to relate the spectral densities of the input $h_y(\omega)$ and of the output $h_g(\omega)$ as follows:

$$h_g(\omega) = Y(\omega) h_y(\omega)$$

where $Y(\omega)$ represents the transfer function, in this case the Fourier Transform described in Section 2. The absolute value of the transfer function is the gain of the filter:

$$G(\omega) = |Y(\omega)|$$

and its power is the squared gain of the filter, which determines how the variance of the input contributes to the variance of the output at different frequencies.

References

- Bracewell, R. (1965). *The Fourier transform and its applications*. McGraw-Hill. (paperback edition: 1999).
- Dagum E.B. (1980). The x-11-arima seasonal adjustment method. Technical report, Statistics Canada.
- Dagum E.B., Bianconcini S. (2015). A new set of asymmetric filters for tracking the short-term trend in real-time. *The annals of applied statistics* 9, 3, 1433–1458.
- Dagum E.B., Chhab N., Chiu K. (1996). Derivation and properties of the X-11-ARIMA and census X11 linear filters. *Journal of official statistics* 12, 4, 329–347.
- Findley D.F. (2005). Some recent developments and directions in seasonal adjustment. *Journal of official statistics* 21, 2, 343–365.
- Findley D.F., Martin D.E.K. (2006). Frequency domain analyses of SEATS and X-11/X-12-ARIMA seasonal adjustment filters for short and moderate-length time series. *Journal of official statistics* 22, 1, 1–34.
- Findley D.F., Soukup R.J. (1999). On the spectrum diagnostics used by X-12-ARIMA to indicate the presence of trading day effects after modelling or adjustment. Technical report, U.S. Census Bureau.
- Grether D.M., Nerlove M. (1970). Some properties of ‘optimal’ seasonal adjustment. *Econometrica* 38, 5, 682–703.
- Grudkowska S. (2015). Jdemetra+ reference manual version 1.1. Technical report, Narodowy Bank Polski.
- Maravall A. (2008a). Singal extraction in ARIMA time series. Technical report, Bank of Spain.
- Maravall A. (2008b). Time series regression with ARIMA noise, missing observations and outliers. Technical report, Bank of Spain.
- Maravall A., Gomez V. (1997). Guide for using the programs TRAMO and SEATS. Technical report, Bank of Spain.
- Maravall A., Sanchez J. (2000). An application of TRAMO/SEATS: Model selection and out-of-sample performance. the swiss cpi series. Technical report, Bank of Spain.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1992). *Numerical Recipes in FORTRAN 77 2nd Ed. : the art of scientific computing*. Cambridge Univ. Pr.
- Shumway R.H., Stoffer D.S. (2011). *Time series analysis and its application 3rd Ed*. Springer.

Publisher

Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress: Statistics Netherlands, Grafimedia
Design: Edenspiekermann

Information

Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order

verkoop@cbs.nl
Fax +31 45 570 62 68
ISSN 1572-0314

© Statistics Netherlands, The Hague/Heerlen 2014.
Reproduction is permitted, provided Statistics Netherlands is quoted as the source