Discussion Paper

Nowcasting using linear time series filters

Frank P. Pijpers

March 2018

Contents

In the production of official statistics it is a common problem that there is a time lag between the date on which collection of source data is finalised and the date on which all administrative and technical processes of quality assurance and of data cleaning are completed. This is true both for survey-based statistics and for register based statistics. This time lag can be sufficiently large that a requirement for timely production of statistics cannot be met without some form of forward extrapolation intended to produce official statistics as they are expected to be at the date of publication. This is commonly referred to as 'nowcasting' of time series. There are a number of ways in which such extrapolation might be achieved, for instance with (only) partially complete data or not fully quality assured data relevant to the time interval between now and $now - timelag$. Another option is to use related leading indicators. This paper focuses on a technique which uses a linear decomposition to separate out trend, seasonal influences, and noise (cf. Perrrucci and Pijpers (2017)) to facilitate a forward extrapolation of the trend and seasonal components, including an estimation of the confidence interval. The method is demonstrated using a time series for numbers of unemployment benefits recipients in the Netherlands, available on the Statistics Netherlands website.

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 of movements of the series throughout the year, with similar intensity in different years. It means that seasonal effects are expected to be predictable with moderate or small uncertainty. Many techniques exist to deal with analysis of time series and extensive descriptions can be found in Harvey (1989), or Durbin and Koopman (2012). Within official statistics the X13-ARIMA and JDemetra software packages incorporate a number of techniques for seasonal adjustment (Caporello and Maravall, 2004; Grudkowska, 2015, 2017). A recent comprehensive overview of issues and methods in the area of official statistics time series is van den Brakel et al. (2015, 2017). The longer term trends may or may not have cyclical behaviour on long time scales but over those time scales the margins of uncertainty are likely to be larger than for the seasonal effects. However, for shorter time scales these variations are, by definition, more coherent, so that extrapolation over time scales of less than a year is feasible. The highest frequency, stochastic, signal in a time series might be amenable to modeling using ARIMA techniques, but is otherwise a source of uncertainty, which ultimately limits the extent over which extrapolation and hence nowcasting is possible.

Viewed in this way, the problem of nowcasting can be approached as an application of seasonal adjustment, with separate distinct extrapolation techniques applied to two of the components in the decomposition (trend and seasonal). The noise component of this decomposition is used to evaluate the confidence intervals around the extrapolated time series which is also the determining factor for the extent to which nowcasting without auxiliary information is feasible. These three components are determined using a linear filtering technique described in Perrrucci and Pijpers (2017), precisely because its linearity is advantageous in the nowcasting.

To demonstrate the performance of this scheme, it is applied to a time series of the number of unemployment benefits in the Netherlands (source: www.cbs.nl/Statline) which is a register-based time series, available monthly from January 1998 onwards.

2 Decomposition

The decomposition of the time series Y, sampled at the discrete times t_i, with regular spacing, is the usual separation into trend+cycle C, seasonal S, and noise components H:

$$Y(t_i) = C(t_i) + S(t_i) + H(t_i) \quad (1)$$

The application to the unemployment benefits data is shown in fig. 2.1 and 2.2. The numbers shown are in 1000's of benefits (: "WW uitkeringen"). The decomposition is performed using

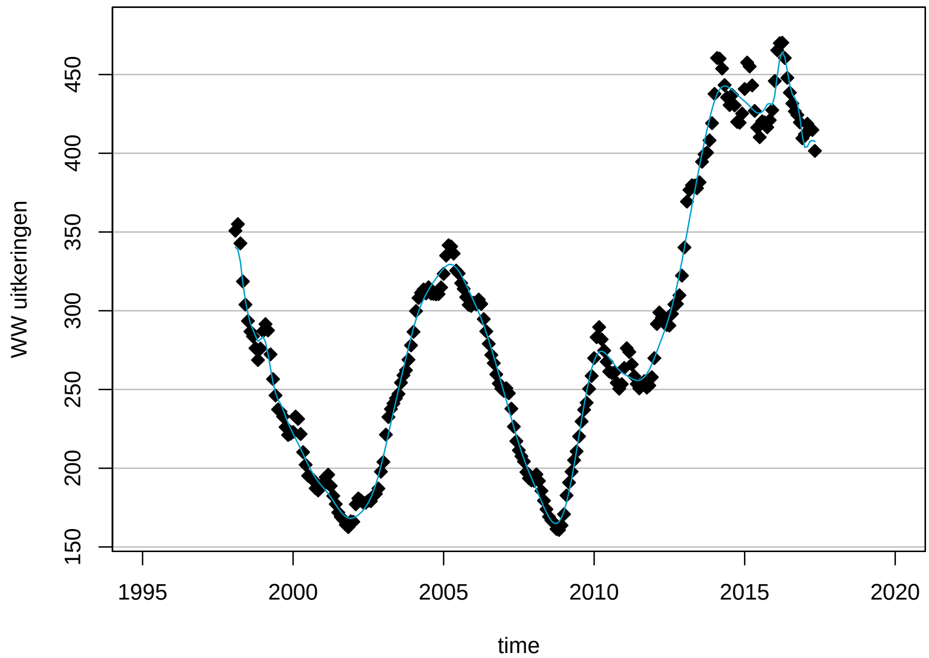

Figure 2.1 The unadjusted data (black symbols) and the trend C (blue line) determined using the linear filter. The numbers of benefits, WW uitkeringen, are expressed in units of one thousand.

the linear filter presented in Perrucci and Pijpers (2017). For convenience the relevant filter weights, and the Fourier transform of these weights are reproduced in the appendix. In fig. 2.2 the data is shown, after removing the trend, ie. $Y - C \equiv S + H$ (black symbols). Also shown is just the seasonal term S (blue line). From this figure it is evident that while there is a clear seasonal influence, the overall pattern is not perfectly constant. For instance, the amplitude of the seasonal variation appears to be systematically larger after 2010, than it was before 2010, perhaps reverting to a smaller amplitude in the most recent year.

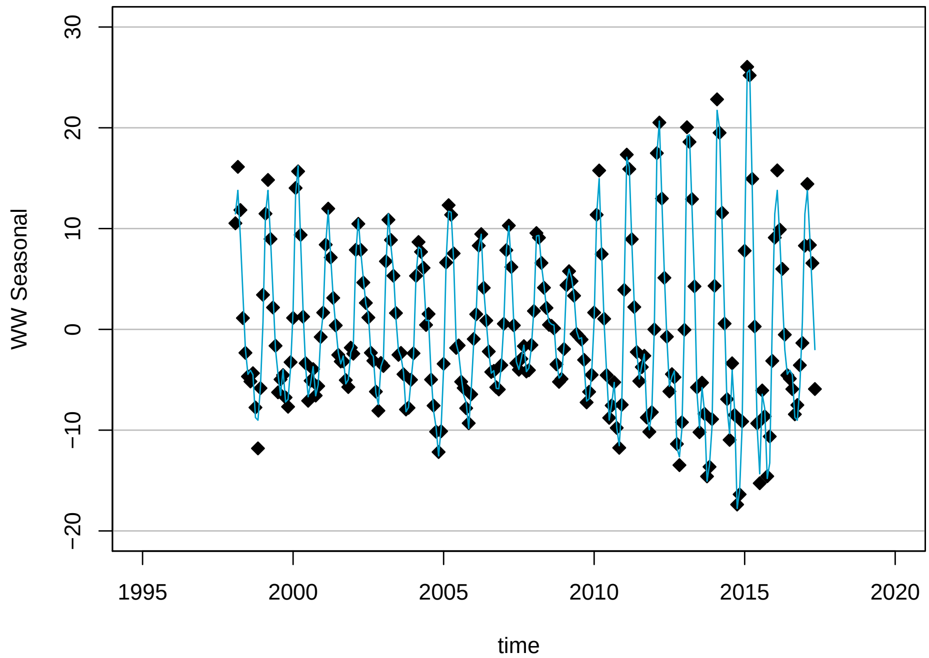

Figure 2.2 The data after removal of the trend+cycle (black symbols) component; ie. the sum of the seasonal and noise components. Also shown is the seasonal component S (blue line) determined using the linear filter. The numbers of benefits, WW uitkeringen, are expressed in units of one thousand.

3 Extrapolation technique

3.1 seasonal term

The linear filtering process described in detail in Perrrucci and Pijpers (2017) and briefly in the appendix, shows that it is in essence taking a moving weighted average of the raw time series, over a window and with weights arranged symmetrically around the time for which this average is calculated. As a result, for the filter properties used here, the seasonal component for the first 18 months of the time series and for the most recent 18 months cannot be determined directly since it would require data belonging to times before the measurements started or data that have not yet been possible to measure. For these subranges, the seasonal term is estimated instead using the procedure outlined below, and this estimation is shown together with the output of the filtering procedure in fig. 2.2.

From fig. 2.2 it is clear that while there is a pattern of intra-year variation, it does not perfectly reproduce from one year to the next. Other data might or might not behave more regularly. For this reason it appears that the most robust approach is to determine the average value of the seasonal component for each calendar month m, and also the r.m.s. for each of the calendar months. If there are M_m calendar months m in the time series with a total number of samples N, the average is:

$$\bar{S}(m) = \frac{1}{M_m} \sum_{k=0}^{M_m - 1} S(t_{m+12k}) \qquad m = 1, ..., 12 \quad 19 \leq (m + 12k) \leq N - 19 \qquad (2)$$

and the r.m.s. is determined correspondingly.

$$\sigma_S(m)^2 = \frac{1}{M_m - 1} \sum_{k=0}^{M_m-1} \left[S(t_{m+12k}) - \bar{S}(m)\right]^2 \qquad m = 1, \ldots, 12 \quad 19 \leq (m + 12k) \leq N - 19 \tag{3}$$

For these summations only those months are used that are not near the edges of the time series (ie. not the first 18 months or the last 18 months), so that there is a direct determination from the linear filter.

Outside of the edges of the time series, and in particular forwards in time, is the epoch for which the extrapolation becomes a proper nowcast or forecast because no data are yet available. In the absence of auxiliary data, the option with arguably the least amount of modeling or additional assumptions, is to continue the same scheme as outlined above for the edges of the time series. That means that for the first (few) months after the end of the measurements, for the seasonal component S the appropriate average value for that month $\bar{S}$ is substituted. It is evident that the confidence intervals around this value will increase in size as one moves forward in time. This is addressed in section 4.

In those parts of the time series where the seasonal component can be determined directly using the linear filters, the noise component is used to determine the confidence intervals for the seasonal component (see section 4). At the edges of the time series the confidence intervals are larger because of the extrapolation backward and forward in time; this is shown in fig. 4.1. In these regions the seasonal component is extrapolated using the above average (eq. (2), and the confidence intervals are determined using the corresponding r.m.s. (eq. (3)). Note that in order to use eq. (3), it is necessary to have more than one instance of every calendar month in the 'internal' part of the time series where the filter completely determines the seasonal component. The implication is that with 18 months at beginning and end of the time series plus a minimum of 24 months in between, the minimally required total length of the time series is 60 months. If fewer months are available, i.e. less than 5 years of data, the extrapolation can still be carried out but it becomes problematic to assess margins of uncertainty. With less than 4 years of data, even the extrapolation itself may become difficult since then for some calendar month or months there is no $\bar{S}(m)$ from Eq. (2). It is therefore not advisable to use this method for nowcasting if less than 4 years of monthly data is available for establishing trends and seasonal behaviour.

Conversely, if the amount of data is large enough, i.e. a time series with a total length of decade or more, one can consider whether it is appropriate for the extrapolation forwards to use a weighting scheme in Eqs. (2) and (3). In this way it would be possible for instance to give more relative weight to more recent data.

$$\begin{aligned}
\bar{S}(m) &= \sum_{k=0}^{M_m-1} w_k(m) S(t_{m+12k}) \qquad m = 1, \ldots, 12 \quad 19 \leq (m + 12k) \leq N - 19 \\
\sigma_S(m)^2 &= \frac{M_m}{M_m - 1} \sum_{k=0}^{M_m-1} w_k(m)^2 \left[S(t_{m+12k}) - \bar{S}(m)\right]^2 \Big/ \sum_{k=0}^{M_m-1} w_k(m)^2 \\
\sum_{k=0}^{M_m-1} w_k(m) &\equiv 1
\end{aligned} \tag{4}$$

An example of a weighting scheme, linear in the index k enumerating the years, would be:

$$w_k(m) = \frac{1}{M_m}\left[1 + \alpha \frac{2k - M_m}{M_m}\right] \qquad 0 \leq \alpha < 1 \quad m = 1, \ldots, 12 \tag{5}$$

The parameter α in (5) can be employed to adjust the weighting scheme. The choice $\alpha = 0$ might be considered a 'default' choice, corresponding to uniform weighting (i.e. Eqs. (2) and (3)). As α increases, the weighting scheme ensures an increasing relative influence of more recent years. In practice some experimentation may be necessary for any given series to establish whether non-uniform weighting is an improvement over uniform weighting and whether there is an optimal choice for α. In the example discussed in this paper only uniform weighting, $\alpha = 0$, is used.

3.2 trend+cycle term

For the same reason that the seasonal term cannot be determined directly in the first and last 18 months of the measured time series, the trend+cycle term also cannot be determined directly in those same months. However, given the extrapolation of the seasonal term, the trend+cycle term can be determined by subtracting the extrapolated seasonal term from the raw data and using Eq. (20) (appendix) to remove the noise. This allows determining the trend term for the first and last 18 months, because raw data is available. Since the confidence intervals for the seasonal term are larger in these regions, so are the confidence intervals for the trend+cycle term, which is addressed in section 4.

Forward in time from the most recent measured data, the trend needs to be extrapolated further for a complete nowcast or forecast of the time series. Contrary to what is the case for the seasonal component there is no guidance for the behaviour of Trend+Cycle using the type of averaging that could be applied to the seasonal term. Instead it makes more sense, given that this component has been constructed to vary slowly, to fit a low order polynomial to a short section of the most recent trend+cycle time series and use this polynomial fit to construct the forward extrapolation.

For a robust extrapolation, as well as a treatment of confidence intervals, it is most convenient to use a set of orthonormal polynomials for the fitting procedure. The confidence intervals around each of the points in the trend+cycle time series can be used to determine confidence intervals for the fitting coefficients. In turn, these are used to determine confidence intervals around the extrapolation of the trend+cycle series.

The purpose at hand is to enable nowcasting over a time lag that is most likely less than a year. Statistics Netherlands as a national statistical agency normally does not do forecasts. However, for national statistical agencies it is very important to detect turning points in trends, i.e. local minima or maxima, at the earliest opportunity. This implies that a linear extrapolation of the trend+cycle term has insufficient degrees of freedom. The next lowest degree polynomial with the same number of even and odd terms has degree 3. Therefore, the fitting function used consists of at most 4 orthonormal polynomials of resp. degree 0, 1, 2, and 3. The advantage of using orthonormal polynomials is that each of the 4 linear coefficients determined by fitting to the end section of the trend+cycle series can be determined statistically independently from the others, and treatment of errors is therefore particularly straightforward since the error covariance matrix for those coefficients is diagonal. By construction, the autocorrelation time scale of the trend+cycle time series is roughly equal to the half-width of the filtering window, i.e. 18 months. It is considered sufficient to determine the fitting coefficients for the polynomial fit to 24 months, the smallest integer number of years with a length larger than the autocorrelation length. The fit of a polynomial of degree 3 (at most) to 24 samples/months of the time series is then used to extrapolate up to 12 months ahead of the leading edge of the measured series. In practice the confidence intervals quickly increase in size, so that going any further forward in time appears to be without merit.

Evidently, the confidence intervals are already larger towards the edges of the time series, because of the previous step, and continue to increase in size the further forward one extrapolates into the nowcast/forecast region.

4 Stochastic component and confidence intervals

While the trend, cycle and seasonal components do have some stochasticity, what is usually termed the noise component in a decomposition, or the stochastic component, is *purely stochastic*: all other components have non-zero expectation values for some or all sampling times, but the noise component does not. The stochastic component of the time series is determined using eq. (19) (see appendix) over that part of the time series where the filtering procedure allows a direct determination. The same equation (19) can also be applied to the edges of the time series, after subtraction of the seasonal term, determined as described by eq. (2). The resulting time series for the (stochastic) residual H is shown in fig. 4.1.

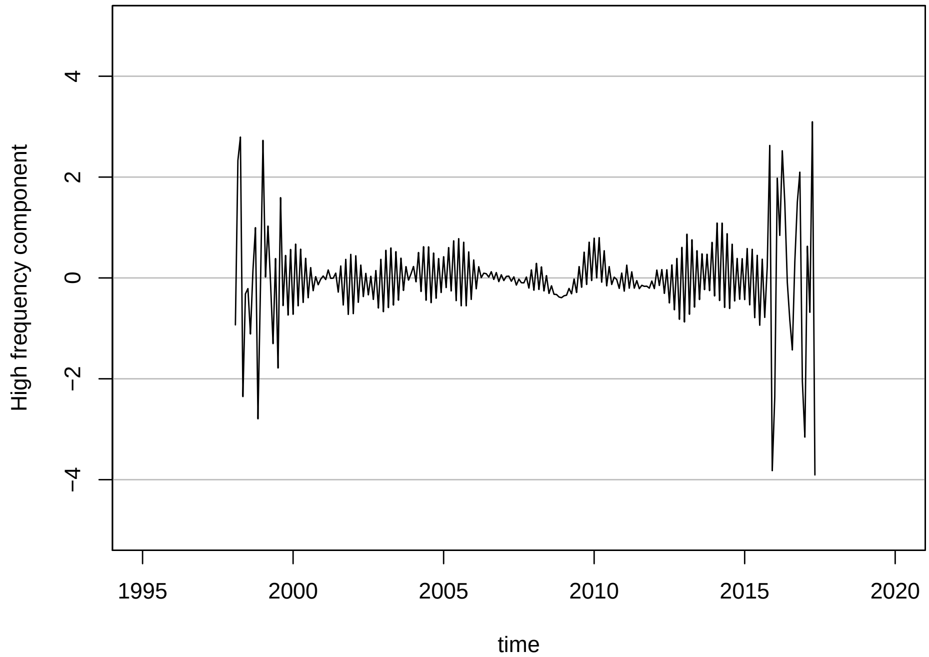

Figure 4.1 The data after removal of the trend+cycle as well as removal of the seasonal component determined using the linear filter. The residual numbers of benefits are the stochastic component H, expressed in units of one thousand.

In fig. 4.1 the effect of the extrapolation can be seen clearly in the behaviour of H at the edges of the time series: in absolute value H reaches larger values by factors of 2 to 4 than in the rest of the time series. In the present case the stochastic term is a relatively small contribution to the overall level of the time series. However, for other time series the trend+cycle C or seasonal term S might be smaller compared to the noise term H. Evidently, the more a time series consists of, or is contaminated by, a stochastic component or noise the more difficult it is to do any kind of nowcasting.

The r.m.s. of H is necessary to enable the calculation of confidence margins, not only for the trend+cycle and seasonal terms, but also for the extrapolations and nowcast. In the Fourier domain the stochastic component occupies the highest frequency band, ie. for the monthly cadence and the filter used the band from 5.25 cycles/year up to 6 cycles/year (see fig. 8.1 in the appendix). The r.m.s. of the noise is equivalent to the power of the noise in that frequency band. In the absence of further detailed knowledge of the process that has generated that noise, the least restrictive assumption for its behaviour is to assume that it is white noise, which is to say that the power in the noise is independent of frequency. This assumption allows a determination of confidence limits for the seasonal and trend+cycle components. The width of the confidence interval for the trend+cycle component is, under this assumption, identical to the width of the confidence interval, ie. the r.m.s., of the noise because the width of the frequency band is the same for the trend+cycle and for the noise. The width of the frequency band for the seasonal component is different from the width of the frequency band for the noise. In the case of the filter shown for the monthly sampled series (fig. 8.1) this bandwidth for the seasonal component is a factor of 6 wider than the bandwidth for the noise component, and so the variance is also a factor of 6 larger. In the case of the filter shown for the quarterly sampled series (fig. 8.2) this bandwidth for the seasonal component is a factor of $2/3$ times the bandwidth for the noise component, and so the variance is equal to $2/3\times$ the variance of the noise component.

Towards the edges of the time series, the above determination of confidence intervals breaks down, because the seasonal component is determined using the scheme described in section 3.1, with the associated larger uncertainties. The r.m.s. determined in section 3.1 can be used directly as a measure of the confidence interval for the seasonal term.

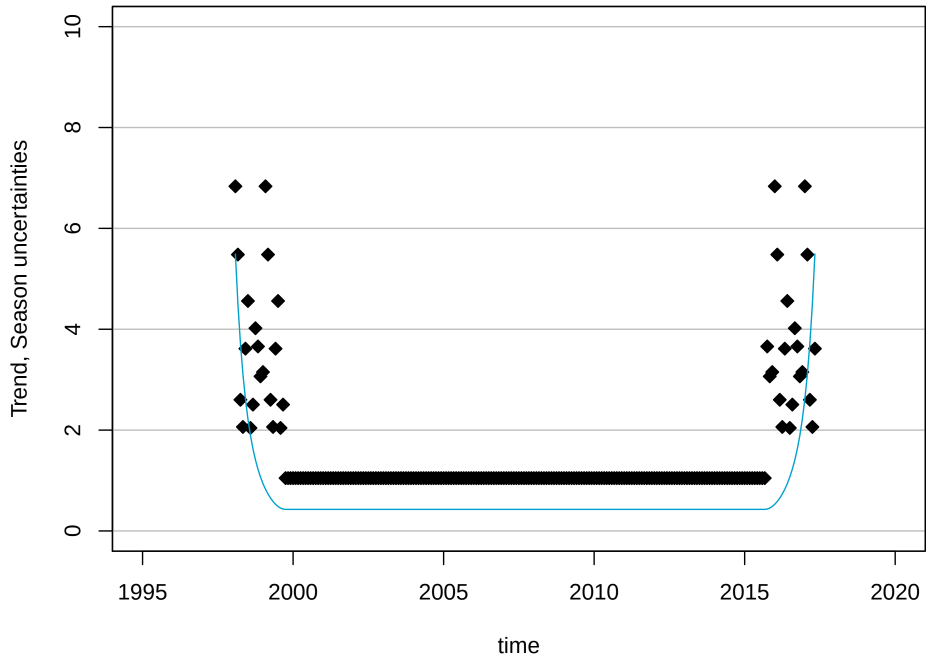

Figure 4.2 The width of the confidence intervals for the seasonal component (black symbols) as well as for the trend+cycle (blue line). The numbers of benefits are expressed in units of one thousand.

For the trend+cycle term the width of the confidence interval must increase gradually as more and more of the filter weights cannot be used, because the required data are missing. The absolute value of the filter weights from the edge of the filter towards the center, for the monthly sampling filter, behave approximately as:

$$|w_{2k}| \approx 0.04 \exp\left[4\left(\frac{20-2k}{19}\right)^2\right] \qquad k = 1, .., 9 \tag{6}$$

The variance for the trend+cycle component gradually increases from its value in the central range of the time series to a value that is typical for the values of the seasonal component at the edges, where the behaviour (6) of the envelope of the weights can be used to model this transition. While this is merely an approximation of the behaviour and other factors influence the uncertainty, it does capture the dominant behaviour. Fig. 4.2 shows the resulting 1σ widths of the confidence intervals for the trend+cycle and seasonal components of the time series. In fig. 4.2 it can be seen that over the first and last 18 months of the time series the dominant source of uncertainty changes over from being the seasonal component to becoming the trend+cycle component.

Outside of the measured time series, in the nowcast/forecast region, the contribution to the width of the confidence interval coming from the seasonal component can be assumed to be similar to the values at the edges of the measured time series. For the trend component, a confidence interval for short intervals outside of the measured time series can be obtained by using the extrapolated polynomial fit and the confidence limits on the fitted coefficients (see appendix). This way of extrapolating the series into the nowcasting regime and the associated uncertainties, results in that quickly the uncertainty coming from the trend+cycle extrapolation is the dominant noise source, and within that it is the highest order term of the fit that eventually dominates. If a cubic polynomial is used for the extrapolation it is the highest order (cubic) term that will dominate, even if initially it is quite small. The consequence is that the uncertainty margin does not increase very much for the first few months into the nowcast region, after which there is a rapid increase which to the eye appears quite sudden.

4.1 other nowcasting methods

Some methods, such as single exponential smoothing (see eg. NIST and SEMATECH (2012)) explicitly assume that there is a well-defined mean to the time series. This can always be eliminated from the problem, so that it is equivalent to a time series with 0 mean. If it is known a-priori that there is a well defined mean, uncertainties from extrapolation will not continue to grow since one can replace the extrapolation at large times with that mean and the uncertainties do not grow beyond the overall series variance: exponential smoothing used as extrapolation does so by construction. In reality time series may not have such a well defined mean. In the framework of exponential smoothing, double or triple smoothing can be carried out to include respectively trend and seasonal terms. In all of these variations of exponential smoothing, the weight of the last measured points decreases exponentially with time which implies that, without adding assumptions about a mean or trend, the uncertainties must increase exponentially with time. For nowcasting applications, ie. relatively short-term forecasts, the difference between uncertainties that increase either polynomially (as is the case here) or exponentially is small and could be in favour of either method, depending on the situation. Structural Time series models (STM) and Kalman filters (cf. Harvey (1989) and van den Brakel et al. (2015)) may make more use of autocorrelation characteristics of the stochastic component. For some timeseries that use of the noise autocorrelation by these methods can enable extending the range in time over which acceptable forecasts can be constructed. It will depend on the particular timeseries and nowcast/forecast needs whether the additional level of complexity of the analysis of those methods is warranted. It is planned to compare various nowcasting methods in terms of the quality of nowcasts, which will be reported on separately.

5 Outlier diagnostic

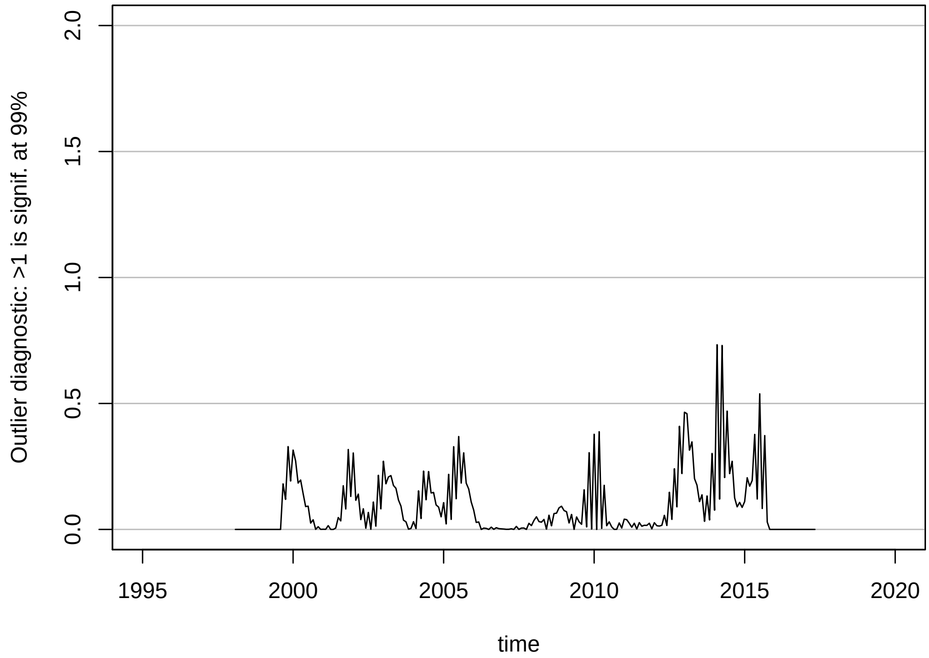

Figure 5.1 The outlier diagnostic evaluated for all points in the measured time series where the trend+cycle, seasonal, and stochastic contributions can be determined directly from the filtering procedure. Outside this range, near the edges of the time series, the diagnostic has no meaningful interpretation. The scale is set so that a value exceeding 1 corresponds the the 99% confidence level that there is an outlier at/near that point

With the time series for the stochastic component in hand (see fig. 4.1) it is possible to calculate the contribution $H(t_i)^2$ at each time t_i to the total variance of the stochastic component, for that range of the time series where a direct determination is possible using the linear filtering. Outside of this range there is no reliable method to determine whether a measurement is an outlier.

The ratio of H^2 at t_i to the total variance can be shown to satisfy a χ^2 distribution with one degree of freedom. The appropriate corresponding scale from its distribution function can then be used to construct a diagnostic function for outliers, which is shown in fig. 5.1 for the time series at hand. The scale is chosen such that a value of 1 corresponds to a 99% confidence level. This means that any peak that exceeds this level has a probability of $< 1\%$ to arise from purely random variations.

Note that the filtering process has the consequence that the variations have a correlation timespan of roughly half the width of the filter. A single outlier could therefore produce several adjacent peaks all exceeding the threshold. The diagnostic is therefore useful for the detection of potential outliers, but has less value for the construction of an automated removal of such outliers. Such techniques are outside the scope of this paper. Other forms of disruption of the measured time series, such as abrupt level or slope changes do produce a signature in this diagnostic, but will display a more complex signature. In such cases a combined examination of trend and stochastic components is required to deal with such issues. A framework for dealing with such breaks, such as eg. caused by transitions in data collection can be found in van den Brakel et al. (2017).

6 The nowcast

6.1 extrapolation and confidence interval

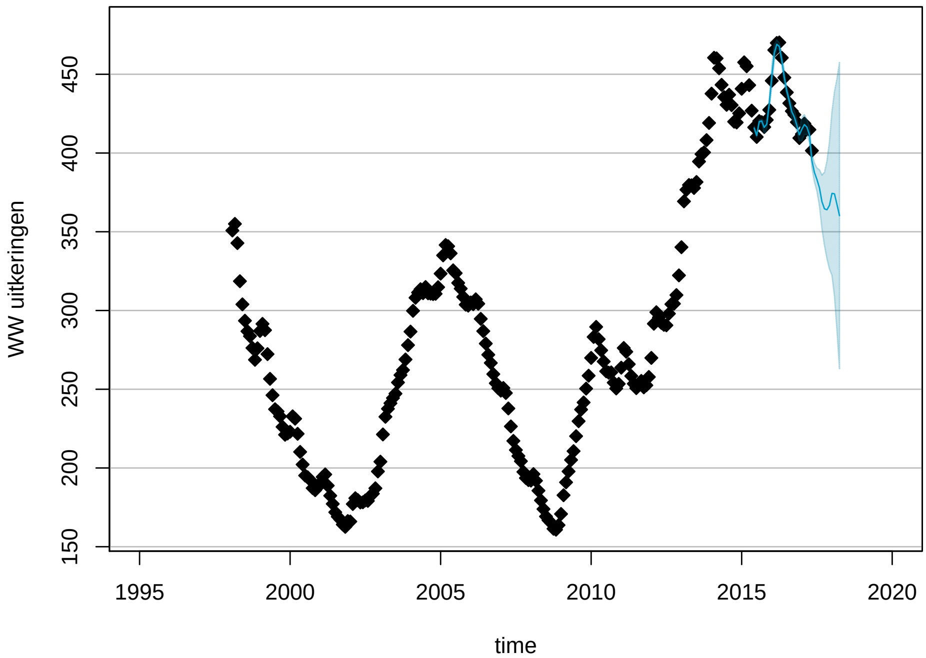

Figure 6.1 The original time series (black symbols) and the sum of the extrapolation of the trend+cycle and seasonal components (blue line) for up to 12 months beyond the last measurement. The area shaded in blue around the extrapolated time series shows the confidence interval. The numbers of benefits are expressed in units of one thousand.

With the decomposition presented in sections 2 and 4 and the extrapolation scheme described in section 3 the time series of unemployment benefits (WW uitkeringen) can be forecast. In fig. 6.1 the resulting fit and extrapolation is shown, together with the measured data. The width of the confidence interval is indicated by the light blue shaded area around the extrapolated line, which has both the trend+cycle and seasonal contributions.

From fig. 6.1 it is seen that the width of the confidence interval increases rapidly. At the last measured point of the time series the confidence interval for this time series is 4.5 thousand. At the nowcast/forecast 12 months ahead of this time, the width of this interval is already 97 thousand, which effectively renders the forecast useless. At the nowcast/forecast 6 months ahead, the width of this interval is 23 thousand, which is of the same order as the amplitude of the seasonal effects (fig. 2.2) and roughly 10% of the total variation in the time series (fig. 6.1). Arguably the extent to which a nowcast provides informative extrapolations, for a time series with this amount of stochastic contributions, is certainly no more than 6 months. At a nowcast/forecast 3 months ahead, the width of this interval is 7.3 thousand which would appear acceptable for many purposes. In principle these uncertainties, and the extent into the future over which informative forecasts can be made, depend on the character of the trend around that point. If, for instance, the most recent epochs of the time series do not show any inflections or extrema in the trend component, it may well be that a nowcast or forecast can be carried out rather further into the future than is the case for this example.

6.2 testing

In order to test whether this extrapolation and its confidence intervals are reliable, a test has been performed where the same time series is truncated, successively at all sampling points between 61 and 219, where the total length of the time series in number of sampling points is 232. For each of these truncated series the entire filtering and fitting process and the extrapolation is repeated, using only the part of the time series up to the truncation point.

Table 6.1 Average Z score and r.m.s. for the first 12 months of the nowcast/forecast of the WW uitkeringen time series

m	$\overline{Z}(m)$	$\sqrt{\overline{Z^2}(m)}$
0	0.01194	0.63799
1	-0.04700	0.97054
2	-0.02995	1.0317
3	-0.02977	1.0929
4	-0.03519	1.1283
5	-0.03803	1.1277
6	-0.03762	1.1079
7	-0.03850	1.0883
8	-0.04114	1.0761
9	-0.04282	1.0597
10	-0.04304	1.0375
11	-0.04237	1.0132

For these truncated time series the actual value is known. This means that the extrapolation Y_e of the time series at point t_{i+k} from the series truncated at t_i can be compared with the actual value y, and a Z-score can be calculated:

$$Z_i(m) \equiv \frac{Y_e(t_{i+m}) - y(t_{i+m})}{\sigma_e(t_{i+m})} \qquad m = 0, \ldots, 11 \;\; i = 61, \ldots, 219 \tag{7}$$

For each k there are therefore $n_t = 159$ Z-scores. The average and r.m.s. of these Z scores is then:

$$\overline{Z}(m) \equiv \frac{1}{n_t} \sum_{i=61}^{61+n_t} Z_i(m)$$
$$\overline{Z^2}(m) \equiv \frac{1}{n_t} \sum_{i=61}^{61+n_t} Z_i(m)^2 \tag{8}$$

A statistically significant non-zero value of $\overline{Z}$ would imply that the scheme is biased and systematically over- or underestimates the time series. For all k the $|\overline{Z}| < 0.05$ and therefore no statistically significant bias is found (see table 6.1). If the value of $\overline{Z^2}$ deviates significantly from 1, this would imply that the confidence intervals are systematically underestimated ($\overline{Z^2} > 1$) or overestimated ($\overline{Z^2} < 1$). For all $m \geq 1$ the value of $\overline{Z^2}$ does not deviate significantly from 1. For $m = 0$ the value is $\sqrt{\overline{Z^2}} \approx 0.64$ which implies that right at the last measured point of the time series, the width of the confidence interval appears to be somewhat overestimated. Over the entire nowcast/forecast range the performance of the extrapolation scheme appears satisfactory.

7 Using auxiliary information

In general nowcasting problems, auxiliary information may be available such as a related time series. This might for instance be a leading indicator, or incomplete data for more recent reporting periods. Since the method described here is linear, some forms of auxiliary information are relatively straightforward to incorporate to produce an improved nowcast.

7.1 single linear equality constraint

One of the simplest forms of auxiliary information is having a single additional time series available, together with a linear relationship between the two time series, ie. between two measured time series Y_1 and Y_2 there is a relationship with the form:

$$aY_1(t) + bY_2(t) = c \tag{9}$$

where a, b and c are known constants. In this case it is convenient to formulate two time series Z_1 and Z_2:

$$\begin{aligned} Z_1(t) &\equiv aY_1(t) + bY_2(t) \\ Z_2(t) &\equiv -bY_1(t) + aY_2(t) \end{aligned} \tag{10}$$

where it is clear that in defining Z_1, use is made of the linear constraint so that $Z_1(t) = c$. The definition of Z_2 is chosen such that it is a linear combination of Y_1 and Y_2 that is orthogonal to the linear combination used in defining Z_1. By defining Z_1 and Z_2 in this way, it is straightforward to demonstrate that there is an inverse of this relationship:

$$\begin{aligned} Y_1(t) &\equiv \frac{1}{a^2+b^2}\left[aZ_1(t) - bZ_2(t)\right] = \frac{1}{a^2+b^2}\left[ac - bZ_2(t)\right] \\ Y_2(t) &\equiv \frac{1}{a^2+b^2}\left[bZ_1(t) + aZ_2(t)\right] = \frac{1}{a^2+b^2}\left[bc + aZ_2(t)\right] \end{aligned} \tag{11}$$

Because of the linearity of the decomposition method, the same relationship that holds between Z_2 and Y_1 or Z_2 and Y_2 also holds between the separate Trend+Cycle, Seasonal, and Noise components of these time series. This means that one can construct the time series Z_2 from the measured series Y_1 and Y_2 using eq. (10), perform the decomposition and nowcast for the single time series Z_2 and reconstruct the decompositions and nowcasts for the Y_1 and Y_2 using eq. (11).

This same method can also be used to deal with mixed frequency data (see also Foroni and Marcellino (2013)) or time series where a partially complete (auxiliary) series is available earlier than a definitive one. At whichever time t_i either y_{1i} or y_{2i} is not available (not measured) the constraint equation $Z_1(t_i) = c$, in combination whichever of the y_{1i}, y_{2i} that *is* available, can be used to nevertheless construct an appropriate value for $Z_2(t_i)$. If for instance one time series extends further forward in time than the other, the available information can be used effectively for a nowcast for both time series.

7.2 multiple linear equality constraints

The case of a single equality constraint can easily be generalised to a situation in which there are multiple auxiliary variables and equality constraints that are linearly independent. If they are not linearly independent, then one or more auxiliary variables and constraints could be removed from the problem without costs in terms of the quality of nowcasts. If there are M variables $Y_1, Y_2, \ldots, Y_M$ and $M - N$ equality constraints, then a set of variables $Z_1, \ldots, Z_M$ can be constructed. Define $\tilde{Y}$ to be the vector containing the M variables Y, and the vector $\tilde{Z}$ the vector of Z variables then:

$$\tilde{Z} = A \cdot \tilde{Y} \tag{12}$$

which is the equivalent of eq. (10). The first $M - N$ rows express each of the $M - N$ equality constraints. The final N rows are constructed to be orthogonal to the first $M - N$ rows, which is always possible. The matrix $M \times M$ matrix A is invertible, given the assumption that all the constraints imposed are linearly independent. This means that:

$$\tilde{Y} = A^{-1} \cdot \tilde{Z} = A^{-1} \cdot \begin{pmatrix} c_1 \\ c_2 \\ .. \\ c_{M-N} \\ Z_{M-N+1}(t) \\ Z_{M-N+2}(t) \\ .. \\ Z_M(t) \end{pmatrix} \tag{13}$$

so that in this case, N time series decompositions are necessary, of the time series $Z_{M-N+1}(t) \ldots Z_M(t)$. The time series for the Y, separately for each decomposition component whenever this is required, are constructed using the inverse relationship (13).

7.3 inequality constraints

In most cases an inequality constraint can be transformed to an equality constraint with the same form as eq. (10), where c is a parameter rather than a simple constant. The same procedure is followed as in the previous sections for single or multiple constraints. The procedures to construct time series decompositions and nowcasts are the same as before, but the parameters c arising from the inequality constraints of course do affect the components for all the Y since there are undetermined degrees of freedom.

A slight variation on this is a constraint where two time series Y_1 and Y_2 are correlated in the sense that their trends are the same, and perhaps also the seasonal components, but not the noise component. This normally means that a solution is sought for which:

$$\sum_i \left[a y_{1i} + y_{2i}\right]^2 \tag{14}$$

is minimised, which is the case if:

$$a = -\frac{\sum_i y_{1i} y_{2i}}{\sum_i y_{1i}^2} \tag{15}$$

This then leads to a form of Eq. (10) where a is given by Eq. (15) and $b = 1$. The trend (and seasonal) components of the time series for the Y are then reconstructed using (11) by setting $c = 0$ and substituting the trend (or seasonal component) of Z_2. This type of quadratic minimisation is one example of adjustment of separate time series, in order to produce mutually consistent series, for which a now classical reference is Denton (1971).

7.4 non-linear constraints

In some nowcasting or time-series problems, there may be non-linear relationships between several measured time series. In many cases it is possible to express this as:

$$\begin{aligned} f_1(Y_1, Y_2, \ldots, Y_M) &= 0 \\ f_2(Y_1, Y_2, \ldots, Y_M) &= 0 \\ &\ldots \\ f_{M-1}(Y_1, \ldots, Y_M) &= 0 \end{aligned} \qquad (16)$$

While it is possible that non-linear techniques provide robust solutions to particular cases, another possible approach is to linearise around a solution $Y_1^*, Y_2^*, \ldots, Y_m^*$ of the system. The Y_i^* are defined as being the (as yet unknown) solution of eq. (16). The linearisation around that then yields:

$$\begin{aligned} \frac{\partial f_1}{\partial Y_1}(Y_1 - Y_1^*) + \frac{\partial f_1}{\partial Y_2}(Y_2 - Y_2^*) + \ldots + \frac{\partial f_1}{\partial Y_M}(Y_M - Y_M^*) &= f_1(Y_1, Y_2, \ldots, Y_M) \\ \frac{\partial f_2}{\partial Y_1}(Y_1 - Y_1^*) + \frac{\partial f_2}{\partial Y_2}(Y_2 - Y_2^*) + \ldots + \frac{\partial f_2}{\partial Y_M}(Y_M - Y_M^*) &= f_2(Y_1, Y_2, \ldots, Y_M) \\ &\ldots \\ \frac{\partial f_{M-1}}{\partial Y_1}(Y_1 - Y_1^*) + \frac{\partial f_{M-1}}{\partial Y_2}(Y_2 - Y_2^*) + \ldots + \frac{\partial f_{M-1}}{\partial Y_M}(Y_M - Y_M^*) &= f_{M-1}(Y_1, Y_2, \ldots, Y_M) \end{aligned} \qquad (17)$$

In this way the problem is reduced to one that has the same form as discussed in section 7.2, because each of the partial derivatives is one element of the matrix A of section 7.2. The final row of that matrix A is again constructed to be orthogonal to all previous rows. The same techniques can then be used to find estimated solutions for each Y_i^*. However, in this case iteration is necessary, because the combination of the Y_i^* thus constructed might not satisfy all the non-linear constraints exactly. This is essentially a multidimensional Newton-Raphson procedure, to be executed for each sample time t. Once a satisfactory solution has been found, ie. the iteration procedure has produced deviations $Y_i - Y_i^*$ smaller than some pre-set value, the decomposition and nowcast can proceed as before.

8 Discussion

This paper presents a method for the nowcasting of time series by extrapolation of the time series without auxiliary information. The procedures to extend this method it to cases where auxiliary information is present are briefly presented as well. An application to a time series without auxiliary information is presented.

In practice the nowcasting problem for the real time series of WW uitkeringen used as an example in this paper does have auxiliary information. There are data available, not presented in this paper, which are direct extracts from (decentralized) government administrations. The issue lies in that this register is continually updated, and in any given month some of these updates are a consequence of measures or procedures which suffer administrative delays. Such updates of an administrative nature produce changes which affect some variables in the register for several months into the past. A time series for such variables, corrected for the changes due to administratively delayed registry updates, will tend to lag by typically 2 to 3 months compared to the time series that are direct database extracts. An application of this method that uses both the direct registration and later corrections to it, i.e. the auxiliary information provided by separate related time series, is beyond the scope of the present work but is to be taken up at a later stage.

This means that the nowcast problem for the particular example of Dutch unemployment benefits (WW uitkeringen) can be framed to refer only to the relative or absolute difference between the time series without and with this accounting for administrative delays. If such a difference implies, for instance, a change in the number of benefits of typically less than 10% of the total number, then the extrapolation uncertainty could be reduced by perhaps the same factor of 10, depending on the relative power in the noise and in the trend+cycle and seasonal components. Over the necessary range for a nowcast of roughly 3 months, to span the lag between the direct 'registration time series' and the time series 'corrected for administrative delays', nowcasts would then have associated confidence intervals in the range of 750.

It should be noted that the seasonal component determined from the filtering (fig 2.2 in section 2) is a combination of calendar effects and 'intrinsic' variation. For many time series in official statistics, the values of the series correlate with the number of working days in a calendar month, which vary from year to year. For the purposes of investigating the causes underlying particular patterns, such calendar effects are of very limited interest. It is therefore likely that for in-depth research into the causes of seasonal variations, one would wish to still correct the seasonal component determined by this filtering, by decorrelating with eg. the number of working days in each calendar month. Such a decorrelation is relatively straightforward to carry out, in particular since this could be done on a time series from which the trend+cycle term has already been removed and a 'clean' seasonal component is available.

The extent to which a nowcast is feasible clearly depends strongly on the power in the stochastic component compared to the other components in the decomposition of the time series, and also on the amount and type of auxiliary information that is available. The methods presented here are implemented in the form of an R module which has no adjustable parameters and is fast. It is therefore very little effort to carry out this analysis on any given time series to asses to what extent the method provides acceptable nowcasts, or whether computationally more complex methods and/or further auxiliary information are required. It is planned to compare various nowcasting methods in terms of the quality of nowcasts, which will be reported on separately.

Appendices

Linear filtering: monthly cadence

In general, filtering in the time domain of a regularly sampled time series can be written in the form of a weighted average:

$$\tilde{y}_j = \sum_{k=-m}^{n} w_k y_{j+k} \tag{18}$$

where the weights w_k are the filter 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 $w_{-k} = w_k$. In Perrrucci and Pijpers (2017) the following filter factors for the even values of k are proposed for a time series with monthly sampling (see table 8.1), for all odd values of k the w_k are identical to 0. The additional property that the weights are zero for all odd values of

Table 8.1 Filter weights for a band-pass filter designed to extract seasonal behaviour from a time series with monthly sampling.

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

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 $\tilde{y}$ 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:

$$h_i = (2\tilde{y}_i - \tilde{y}_i - \tilde{y}_{i+1})/4 \tag{19}$$

In essence this is because this scheme can be seen to be equivalent to an additional high-pass filtering step. The *trend+cycle* time series c (without high frequency contributions) is obtained from:

$$c_i = (2\tilde{y}_i + \tilde{y}_i + \tilde{y}_{i+1})/4 \tag{20}$$

which is a low-pass filter. With the weights of Table 8.1 it is clear that after 18 months any seasonal adjusted time series $\tilde{y}$ will not change, when using this filter alone. Figure 8.1 shows the response function of the filter in the frequency domain: it is designed to block signal with frequencies below about 0.75 cycles/year and also signal with frequencies above about 5.25 cycles/year.

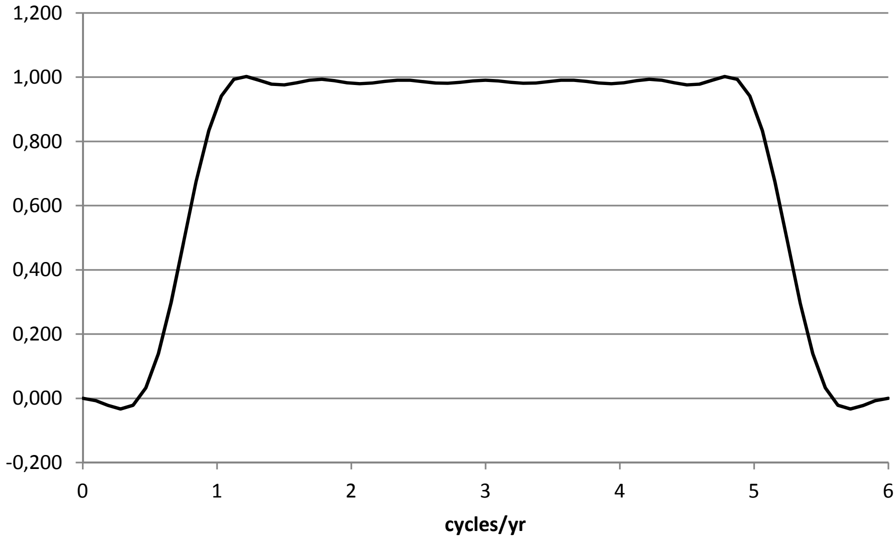

Figure 8.1 The band-pass filter designed for seasonal adjustment, monthly sampling cadence. A filter value of 1 implies that all signal at these frequencies is passed through, whereas a value of 0 implies all signal is blocked.

The transition between passing or blocking signal could be made sharper, but this would have the cost of having non-zero weights w_k for larger values of k. That would mean that at the edges of the measured time series there would be larger sections where no seasonal adjustment could be done because data outside of the measured series would be required. The chosen set is in this sense a compromise.

Linear filtering: quarterly cadence

The reasoning applied to produce the filter factors for a time series that is sampled monthly can be extended to cover other cases such as sampling of once per quarter. However, the practical use of filtering is more limited. There are two main reasons for this.

Firstly, the Nyquist frequency for quarterly sampled time series is lower than for monthly sampled series. Where for time series sampled monthly there is access to signal with frequencies up to 6 cycles/year, for quarterly sampled time series this upper limit is 2 cycles/year. Periodic signal that is part of a seasonal pattern with frequencies between 2 and 6 cycles/year simply cannot be measured with quarterly sampled series.

Secondly, to achieve a similar sharpness of the transition between blocking and passing signal as is obtained for the monthly sampled series, the weights w_k are non-zero over a larger range of k. As can be seen from the table 8.2, the maximum k is now 34 rather than 18 samples. Combined with the fact that the spacing between the quarterly samples is 3 months rather than 1 month, the full width of the time window with non-zero weights is now 204 months $=17$ years, rather than the 3 years for monthly sampled series.

The implication is that while not impossible in principle, performing seasonal adjustment with time series that are sampled at a cadence of once per quarter requires full lengths of time series that are only rarely available in official statistics, and have a time lag of 8.5 years which is not acceptable in most applications where official statistics are used as part of the evidence base for government policy.

Table 8.2 Filter weights for a band-pass filter designed to extract seasonal behaviour from a time series with quarterly sampling.

k	$w_k = w_{-k}$
0	0.2466000
2	-0.2223010
4	0.1523522
6	-0.0680700
8	0
10	0.0339000
12	-0.0345890
14	0.0172851
16	0
18	-0.0069606
20	0.0044180
22	0.0000731
24	0
26	-0.0024583
28	0.0041327
30	-0.0030270
32	0
34	0.0019446

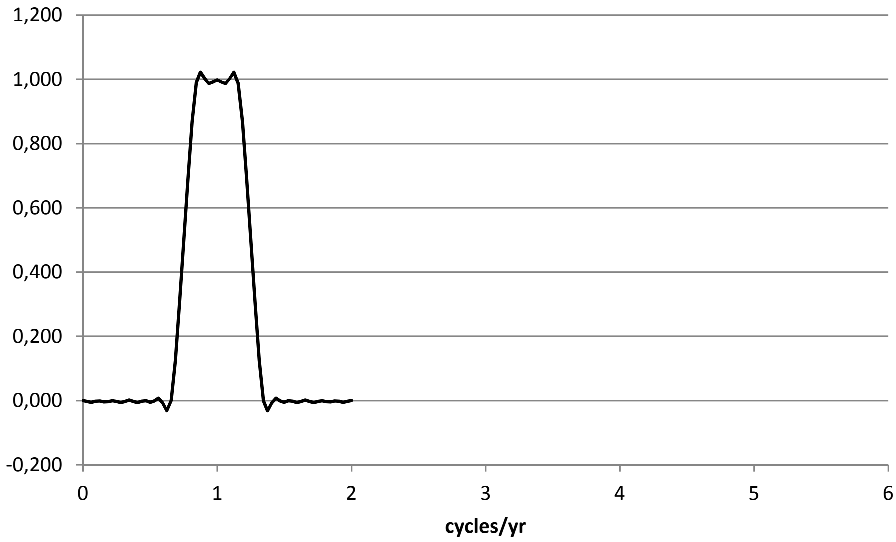

Figure 8.2 The band-pass filter designed for seasonal adjustment, quarterly sampling cadence. A filter value of 1 implies that all signal at these frequencies is passed through, whereas a value of 0 implies all signal is blocked. The region between 2 and 6 cycles/year is inaccessible.

orthonormal polynomials

A set of functions $f_k(x)$ $k \in \{1, \ldots, K\}$ on a finite interval in x of $[a, b]$ that have the property that:

$$0 = \int_a^b f_k(x) f_{k'}(x) dx \quad k \neq k' \tag{21}$$

are orthogonal. There are a number of ways in which they can be defined. All such sets have the convenient property that if they are used for the purposes of fitting to measured data, the fitting process of each orthogonal function can be carried out completely independently of all other components. This also implies that adding more or fewer of such orthogonal functions when fitting to data, will not alter the values of any fitting coefficients already obtained.

The discrete equivalent of the above continuous description is:

$$0 = \sum_{i=1}^{N} f_k(x_i) f_{k'}(x_i) \quad k \neq k' \tag{22}$$

In addition one can normalise each function such that:

$$1 = \sum_{i=1}^{N} [f_k(x_i)]^2 \tag{23}$$

which can be a convenient property in terms of stability of fit. For convenience the moments of the set of values x_i are defined as follows:

$$\begin{aligned} \overline{x} &= \frac{1}{N}\sum_i x_i \\ s^2 &= \frac{1}{N}\sum_i (x_i - \overline{x})^2 \\ S &= \frac{1}{N}\sum_i (x_i - \overline{x})^3 \\ K &= \frac{1}{N}\sum_i (x_i - \overline{x})^4 \\ S_p &= \frac{1}{N}\sum_i (x_i - \overline{x})^p \quad p \geq 5 \end{aligned} \tag{24}$$

In order to obtain normalised functions from these f_k the most straightforward route is to evaluate the sums in (23) and divide each f_k by the square root of the appropriate sum. For the first two polynomials the evaluation is trivial, and the normalised versions of the functions are:

$$\begin{aligned} f_0(x) &= 1 \\ f_1(x) &= \frac{x - \overline{x}}{\sqrt{s^2}} \end{aligned} \tag{25}$$

For f_2 the summation (23) evaluates to $K - (s^2)^2 - S^2/s^2$ so that the normalised equivalent of f_2 is:

$$f_2(x) = \frac{(x - \overline{x})^2 - S/s^2\,(x - \overline{x}) - s^2}{\sqrt{K - (s^2)^2 - S^2/s^2}} \tag{26}$$

The next orthogonal polynomial has degree 3 and has the form:

$$f_3(x) = (x - \overline{x})^3 - S + \beta(x - \overline{x}) + \gamma\left[(x - \overline{x})^2 - s^2\right] \tag{27}$$

in which the values of β and γ are:

$$\begin{aligned} \beta &= \frac{-K\left[K - \left(s^2\right)^2\right] + S\left[S_5 - s^2 S\right]}{s^2\left[K - \left(s^2\right)^2\right] - S^2} \\ \gamma &= \frac{\left[K + (s^2)^2\right]S - s^2 S_5}{s^2\left[K - \left(s^2\right)^2\right] - S^2} \end{aligned} \tag{28}$$

For a perfectly regular spacing of the x_i the odd moments S and S_5 both reduce to 0, so that the expressions for β and γ will simplify to:

$$\begin{aligned} \beta &= -\frac{K}{s^2} \\ \gamma &= 0 \end{aligned} \tag{29}$$

For f_3 the summation (23) evaluates to:

$$\begin{aligned} \frac{1}{N}\sum_i [f_3(x_i)]^2 &= S_6 - S^2 + \beta\left[2K + \beta s^2\right] + \gamma\left[2S_5 - 2s^2 S + 2\beta S + \gamma K\right] \\ &= S_6 - \frac{K^2}{s^2} - S^2 + 2\gamma\left[S_5 - s^2 S - \frac{KS}{s^2}\right] + \gamma^2 \frac{Ks^2 - (s^2)^3 - S^2}{s^2} \\ &= S_6 - \frac{K^2}{s^2} - S^2 - \gamma^2 \frac{Ks^2 - (s^2)^3 - S^2}{s^2} \\ &= S_6 - \frac{K^2}{s^2} - S^2 - \frac{\left\{\left[K + (s^2)^2\right] S - s^2 S_5\right\}^2}{(s^2)^2\left[K - (s^2)^2\right] - s^2 S^2} \end{aligned} \tag{30}$$

Therefore dividing f_3 by the square root of this factor results in a normalised version of f_3. For a perfectly regular spacing, or a spacing which is symmetric around $\overline{x}$, the odd moments $S = S_5 = 0$ from which it follows that $\gamma = 0$ so that equation (30) reduces to:

$$\frac{1}{N}\sum_i [f_3(x_i)]^2 = S_6 - \frac{K^2}{s^2} \tag{31}$$

Least squares fitting of a set of values y_i with errors with standard deviation σ at the sampling points x_i, where the errors are uncorrelated between the different sampling points, using these functions as base set leads to:

$$\begin{aligned} a_k &= \frac{1}{N}\sum_i y_i f_k(x_i) \\ cov(a_k, a_{k'}) &= \sum_i \sigma^2 f_k(x_i) f_{k'}(x_i) \quad (because\ cov(y_i, y_{i'}) = 0\ \forall\ i \neq i') \\ &= \delta_{kk'}\frac{\sigma^2}{N} \end{aligned} \tag{32}$$

This is easily generalised to the case of heteroscedastic (uncorrelated) errors with unequal standard deviations σ_i at the sampling points.

References

Caporello, G. and A. Maravall (2004). Program TSW, Tramo Seats Windows, revised reference manual. Technical report, Banco de Espana, Madrid.

Denton, F. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *J.Am.Stat.Assoc. 66*, 99--102.

Durbin, J. and S. Koopman (2012). *Time series analysis by state space methods, 2^{nd} Ed.* Oxford Univ. Pr.

Foroni, C. and M. Marcellino (2013). A survey of econometric methods for mixed-frequency data. Technical report, Norges bank, WP 2013/06.

Grudkowska, S. (2015). JDemetra+ reference manual version 1.1. Technical report, Narodowy Bank Polski.

Grudkowska, S. (2017). JDemetra+ reference manual version 2.2. Technical report, Narodowy Bank Polski.

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge Univ. Pr.

NIST and SEMATECH (2012). *e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/*. NIST, U.S. Dept. of Commerce.

Perrrucci, E. and F. Pijpers (2017). Filtering in the Fourier domain: a new set of filters for seasonal adjustment of time series and its evaluation. Statistics Netherlands. discussion paper.

van den Brakel, J., G. Griffiths, T. Surzhina, P. Wise, J. Blanchard, X. M. Zhang, and O. Honchar (2017). A framework for measuring the impact of transitions in official statistics. Technical report, Australian Bureau of Statistics.

van den Brakel, J., S. Krieg, P. Ouwehand, F. v. d. Pol, F. v. Ruth, M. Smeets, and P. Verbiest (2015). Data related issues with composite indicators: Availability, frequency and adjustment techniques. Statistics Netherlands. discussion paper.

van den Brakel, J., S. Krieg, P. Ouwehand, F. v. d. Pol, F. v. Ruth, M. Smeets, and P. Verbiest (2017). Data related issues with composite indicators: Availability, frequency and adjustment techniques. In G. Mazzi and A. Ozyildirim (Eds.), *UN handbook on cyclical composite indicators*.

Colophon

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