Alain QUARTIER-LA-TENTE

Real-time detection of turning points with linear filters

Banque Nationale de Belgique

Cellule Recherche et Développement

Maître de stage : Jean PALATE

08/06/2020 - 14/08/2020

Contents

1 Abstract

Abstract

In the business cycle analysis, estimates are usually derived from moving average techniques. When it comes to real-time estimates, classic asymmetric moving average introduce revisions as new observations are available and also delays in the detecting turning points. This paper describes and compare different approaches to build asymmetric filters: local polynomials filters, methods based on an optimization of filters properties (Fidelity-Smoothness-Timeliness, FST, approach and a data-dependent filter) and filters based on Reproducing Kernel Hilbert Space. It also describes how local polynomials filters can be extended to include a timeliness criteria to minimize phase shift.

The paper emphasizes that constraining asymmetric filters to preserve linear and constant trends is enough to reduce revision error and time lag. The more future observations are available, the more the different methods produce similar results. Finally, the FST approach gives promising results to build asymmetric moving averages.

2 Introduction

With the COVID-19 crisis, one of the major questions on the economy was when it will collapse and when it will recover. It illustrates that business cycle analysis, and in particular the early detection of turning points in a series, is a major topic in the analysis of economic outlook.

Moving averages, or linear filters, are ubiquitous in business cycle extraction and seasonal adjustment methods[1]. For example, the X-12ARIMA seasonal adjustment method uses Henderson moving averages and composite moving averages to estimate the main components of a time series, while TRAMO-SEATS uses Wiener-Kolmogorov filters. Symmetric filters are applied to the center of the series, but when it comes to estimate the most recent points, all of these methods must rely on asymmetric filters. For example, X-12ARIMA or TRAMO-SEATS apply symmetrical averages to the forecasts obtained from an ARIMA model of the series. However, since the predicted values are linear combinations of past values, it consists in applying to use methods do use skewed moving averages at the end of the series.

If the classic asymmetric moving averages have good properties regarding the size of future revisions induces by the process[2], they create phase shifts that impact the real-time estimation of turning points, introducing time delay in the detection.

This report aims to describe and compare the recent approaches around trend-cycle extraction and asymmetric filters: local polynomial filters, methods based on an optimization of filters properties and filters based on Reproducing Kernel Hilbert Space. Due to the link between seasonal adjustment and trend-cycle extraction (section 3), we here focus on non-parametric methods that could be included in X-12ARIMA. After a description of the general properties of a linear filter (section 4), we described the linear filters methods developed by Proietti and Luati (2008), Grun-Rehomme, Guggemos, and Ladiray (2018), Wildi and McElroy (2019) and Dagum and Bianconcini (2008) (sections 5 to 8).

---

[1]A moving average is a statistical method that consists in applying a rolling weighted mean to a times series: for each date it computes a weighted mean of p past points and q future points where $p, q \geq 0$ depends on the moving average.

[2]See for example Pierce (1980).

3 From seasonal adjustment to trend-cycle estimation

More and more infra-annual statistics are produced, especially by national institutes, to analyze the short-term evolution of economies. It is for example the case of the gross domestic product (GDP), unemployment rate, household consumption of goods and industrial production indices. However, most of those time series are affected by seasonal and trading days effects. A seasonal effect is an effect that occurs in the same calendar month with similar magnitude and direction from year to year. For instance, automobile production is usually lower during summer, due to holidays, and chocolate sales are usually higher in December, due to Christmas. Trading days effect appears when a time series is affected by calendar month's weekday composition. For example, retail sales are usually higher on Saturday, thus it is likely that they will be higher in months with a surplus of weekend days.

Seasonal and trading days effects can hamper the analysis of infra-annual movements of a time series or the spatial comparison. This is why time series are often seasonally and trading days adjusted, where seasonal adjustment is the process of removing the effects of seasonal and trading day fluctuations.

To perform seasonal adjustment, most of the algorithm decompose the data in several unknown component: trend-cycle, seasonal and irregular component. In X-12ARIMA and TRAMO-SEATS (the most popular seasonal adjustment methods), prior to the decomposition, the initial series is pre-adjustment from deterministic effects (outliers, calendar effects). Thus, the estimation of trend-cycle component is technically linked to the seasonal component.

Moreover, since the trend-cycle extraction methods are applied to seasonally adjusted data, both problems cannot be separated. This link also explains that in this paper, all the methods used are implemented in the core libraries of JDemetra+ [3], the seasonal adjustment software recommended by Eurostat. An **ℝ** interface is implemented in the package `rjdfilters`[4] was developed during this internship.

In this paper we focus methods that could be implemented in X-12ARIMA. To maintain consistency with the non-parametric approach of X-12ARIMA, we focus on non-parametric extraction methods. That's why the filters of the approach of Wildi and McElroy (2019) (section 7) are not used in this report.

Furthermore, for simplicity the data used is seasonally adjusted: the study should be extended to see the impact on the overall seasonal adjustment process.

---

[3] https://github.com/jdemetra/jdemetra-core.
[4] https://github.com/palatej/rjdfilters.

4 Moving average and filters

Lots of papers describe the definition and the properties of moving average and linear filters (see for example Ladiray (2018)). In this section we summarize some of the main results.

Let p et f two integers, a moving average M_θ or M is defined by a set of coefficients $\theta = (\theta_{-p}, \ldots, \theta_f)'$ such as for all time series X_t:

$$M_\theta(X_t) = \sum_{k=-p}^{+f} \theta_k X_{t+k}$$

- $p + f + 1$ is called the *moving average order*.

- When $p = f$ the moving average is said to be *centered*. If we also have $\forall k : \theta_{-k} = \theta_k$, the moving average M_θ is said to be *symmetric*. In this case, the quantity $h = p = f$ is called the *bandwidth*.

4.1 Gain and phase shift functions

Let $X_t = e^{-i\omega t}$. The result of the moving average M_θ in X_t is:

$$Y_t = M_\theta X_t = \sum_{k=-p}^{+f} \theta_k e^{-i\omega(t+k)} = \left(\sum_{k=-p}^{+f} \theta_k e^{-i\omega k} \right) \cdot X_t.$$

The function $\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{-i\omega k}$ is called the *transfer function* or *frequency response function*[5]. It can be rewritten as:

$$\Gamma_\theta(\omega) = G_\theta(\omega) e^{-i\Phi_\theta(\omega)}$$

where $G_\theta(\omega) = |\Gamma_\theta(\omega)|$ is the *gain* or *amplitude* function and $\Phi_\theta(\omega)$ is the *phase shift* or *time shift* function[6]. For all symmetric moving average we have $\Phi_\theta(\omega) \equiv 0 \pmod{\pi}$.

To sum up, applying a moving average to a harmonic times series affects it in in two different ways:

- by multiplying it by an amplitude coefficient $G_\theta(\omega)$;

- by "shifting" it in time by $\Phi_\theta(\omega)/\omega$, which directly affects the detection of turning points[7].

---

[5]The frequency response function can equivalently be defined as $\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{i\omega k}$ or $\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{2\pi i \omega k}$.

[6]This function is sometimes represented as $\phi_\theta(\omega) = \frac{\Phi_\theta(\omega)}{\omega}$ to mesure the phase shift in number of periods.

[7]When $\Phi_\theta(\omega)/\omega > 0$ the time shift is positive: a turning point is detected with delay.

Example: with $M_{\theta_0}X_t = \frac{1}{2}X_{t-1} + \frac{1}{2}X_t$ we have:

$$\Gamma_{\theta_0}(\omega) = \frac{1}{2} + \frac{1}{2}e^{-i\omega} = |\cos(\omega/2)|e^{-i\frac{\omega}{2}}$$

The figure 1 illustrates the gain and the phase shift for $\omega = \pi/2$ and $X_t = \sin(\omega t)$.
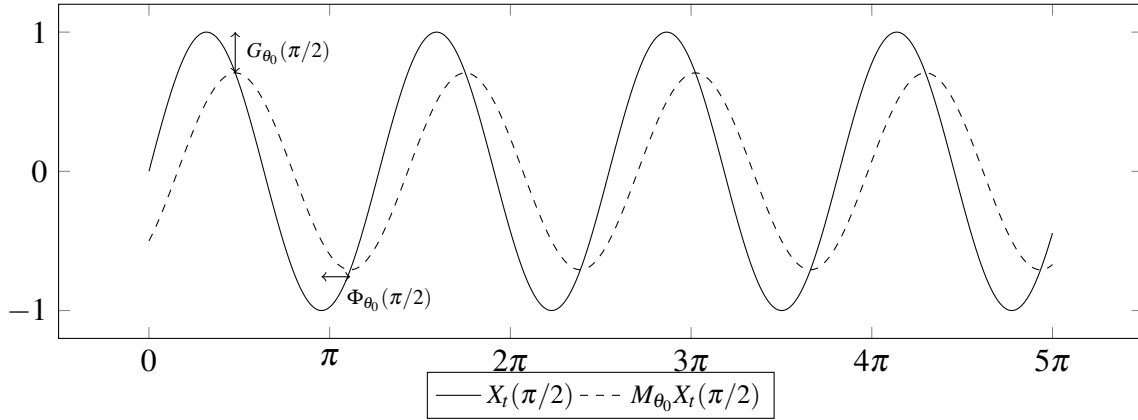
Figure 1: Smoothing of the time series $X_t = \sin(\omega t)$ by the moving average $M_{\theta_0}X_t = \frac{1}{2}X_{t-1} + \frac{1}{2}X_t$ for $\omega = \pi/2$.

4.2 Desirable properties of a moving average

The moving average are often constructed under some specific constraints. In the report we will focus on two constraints:

- the preservation of certain kind of trends;

- the variance reduction.

4.2.1 Trend preservation

It is often desirable for a moving average to conserve certain kind of trends. A moving average M_θ conserve a function of the time $f(t)$ if $\forall t : M_\theta f(t) = f(t)$.

We have the following properties for the moving average M_θ:

- To conserve a constant series $X_t = a$ we need

$$\forall t : M_\theta(X_t) = \sum_{k=-p}^{+f} \theta_k X_{t+k} = \sum_{k=-p}^{+f} \theta_k a = a \sum_{k=-p}^{+f} \theta_k = a$$

the sum of the coefficients of the moving average $\sum_{k=-p}^{+f} \theta_k$ must then be equal to 1.

5

- To conserve a linear trend $X_t = at + b$ we need:

$$\forall t : M_\theta(X_t) = \sum_{k=-p}^{+f} \theta_k X_{t+k} = \sum_{k=-p}^{+f} \theta_k [a(t+k)+b] = at \sum_{k=-p}^{+f} k\theta_k + b \sum_{k=-p}^{+f} \theta_k = at + b$$

which is equivalent to:

$$\sum_{k=-p}^{+f} \theta_k = 1 \quad \text{and} \quad \sum_{k=-p}^{+f} k\theta_k = 0$$

- In general, it can be shown that M_θ conserve a polynomial of degree d if and only if:

$$\sum_{k=-p}^{+f} \theta_k = 1 \text{ and } \forall j \in [\![1,d]\!] : \sum_{k=-p}^{+f} k^j \theta_k = 0$$

- If M_θ is symmetric ($p = f$ and $\theta_{-k} = \theta_k$) and conserves polynomial of degree $2d$ then it also conserves polynomial of degree $2d + 1$.

4.2.2 Variance reduction

All time series are affected by noise that can blur the signal extraction. Hence, we seek to reduce the variance of the noise. The sum of the squares of the coefficients $\sum_{k=-p}^{+f} \theta_k^2$ is the *variance reduction* ratio.

Indeed, let $\{\varepsilon_t\}$ a sequence of independent random variables with $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{V}[\varepsilon_t] = \sigma^2$.

$$\mathbb{V}[M_\theta \varepsilon_t] = \mathbb{V}\left[\sum_{k=-p}^{+f} \theta_k \varepsilon_{t+k}\right] = \sum_{k=-p}^{+f} \theta_k^2 \mathbb{V}[\varepsilon_{t+k}] = \sigma^2 \sum_{k=-p}^{+f} \theta_k^2$$

4.3   Real-time estimation and asymmetric moving average

For symmetric filters, the phase shift function is equal to zero (modulo π). Therefore, there is no delay in any frequency: that's why they are preferred to the asymmetric ones. However, they cannot be used in the beginning and in the end of the time series because no past/future value can be used. Thus, for real-time estimation, it is needed to build asymmetric moving average that approximate the symmetric moving average.

The approximation is summarized by quality indicators. In this paper we focus on the ones defined by Grun-Rehomme, Guggemos, and Ladiray (2018) and Wildi and McElroy (2019) to build the asymmetric filters.

Grun-Rehomme, Guggemos, and Ladiray (2018) propose a general approach to derive linear filters, based on an optimization problem of three criteria: *Fidelity* (F_g, noise reduction),

Smoothness (S_g) and *Timeliness* (T_g, phase shift between input and ouput signals). See section 6 for more details.

Wildi and McElroy (2019) propose an approach based on the decomposition of the mean squared error between the symmetric and the asymmetric filter in four quantities: *Accuracy* (A_w), *Timeliness* (T_w), *Smoothness* (S_w) and *Residual* (R_w). See section 7 for more details.

All the indicators are summarized in table 1.

Sigle	Description	Formula
b_c	Constant bias	$\sum_{k=-p}^{+f} \theta_k - 1$
b_l	Linear bias	$\sum_{k=-p}^{+f} k\theta_k$
b_q	Quadratic bias	$\sum_{k=-p}^{+f} k^2\theta_k$
F_g	Variance reduction / Fidelity (Guggemos)	$\sum_{k=-p}^{+f} \theta_k^2$
S_g	Smoothness (Guggemos)	$\sum_j (\nabla^3 \theta_j)^2$
T_g	Timeliness (Guggemos)	$\int_0^{2\pi/12} \rho_\theta(\omega)\sin(\varphi_\theta(\omega))^2\,d\omega$
A_w	Accuracy (Wildi)	$2\int_0^{2\pi/12} (\rho_s(\omega) - \rho_\theta(\omega))^2\, h_{RW}(\omega)\,d\omega$
T_w	Timeliness (Wildi)	$8\int_0^{2\pi/12} \rho_s(\omega)\rho_\theta(\omega)\sin^2\left(\frac{\varphi_s(\omega)-\varphi_\theta(\omega)}{2}\right) h_{RW}(\omega)\,d\omega$
S_w	Smoothness (Wildi)	$2\int_{2\pi/12}^{\pi} (\rho_s(\omega) - \rho_\theta(\omega))^2\, h_{RW}(\omega)\,d\omega$
R_w	Residual (Wildi)	$8\int_{2\pi/12}^{\pi} \rho_s(\omega)\rho_\theta(\omega)\sin^2\left(\frac{\varphi_s(\omega)-\varphi_\theta(\omega)}{2}\right) h_{RW}(\omega)\,d\omega$

Table 1: Criteria used to check the quality of a linear filter defined by its coefficients $\theta = (\theta_k)_{-p \leq k \leq f}$ and its gain and phase shift function, ρ_θ and φ_θ.
Note: X_g criteria are derived from Grun-Rehomme, Guggemos, and Ladiray (2018) and X_w criteria from Wildi and McElroy (2019).
ρ_s and φ_s represent the gain and phase shift function of the Henderson symmetric filter.
h_{RW} is the spectral density of a random walk: $h_{RW}(\omega) = \frac{1}{2(1-\cos(\omega))}$.

5   Local polynomial filters

In this section we detail the filters that arise from fitting a local polynomial to our time series, as described by Proietti and Luati (2008).

We assume that our time series y_t can be decomposed as

$$y_t = \mu_t + \varepsilon_t$$

where μ_t is the signal (trend) and $\varepsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$ is the noise[8]. We assume that μ_t can be locally approximated by a polynomial of degree d of the time t between y_t and the neighboring

---

[8]The time series is therefore seasonally adjusted.

observations $\left(y_{t+j}\right)_{j\in[\![-h,h]\!]}$. Then $\mu_t \simeq m_t$ with:

$$\forall j \in [\![-h,h]\!] : y_{t+j} = m_{t+j} + \varepsilon_{t+j}, \quad m_{t+j} = \sum_{i=0}^{d} \beta_i j^i$$

This signal extraction problem is then equivalent to the estimation of $m_t = \beta_0$. In matrix notation we can write:

$$
\underbrace{\begin{pmatrix} y_{t-h} \\ y_{t-(h-1)} \\ \vdots \\ y_t \\ \vdots \\ y_{t+(h-1)} \\ y_{t+h} \end{pmatrix}}_{y} = \underbrace{\begin{pmatrix} 1 & -h & h^2 & \cdots & (-h)^d \\ 1 & -(h-1) & (h-1)^2 & \cdots & (-(h-1))^d \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & h-1 & (h-1)^2 & \cdots & (h-1)^d \\ 1 & h & h^2 & \cdots & h^d \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \vdots \\ \beta_d \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \varepsilon_{t-h} \\ \varepsilon_{t-(h-1)} \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_{t+(h-1)} \\ \varepsilon_{t+h} \end{pmatrix}}_{\varepsilon}
$$

Two parameters are crucial in determining the accuracy of the approximation:

- the degree d of the polynomial;

- the number of neighbors $H = 2h+1$ (or the *bandwidth h*).

In order to estimate β we need $H \geq d+1$ and the estimation is done by the weighted least squares (WLS), which consists of minimizing the following objective function:

$$S(\hat{\beta}_0, \ldots, \hat{\beta}_d) = \sum_{j=-h}^{h} \kappa_j (y_{t+j} - \hat{\beta}_0 - \hat{\beta}_1 j - \cdots - \hat{\beta}_d j^d)^2$$

where κ_j is a set of weights called *kernel*. We have $\kappa_j \geq 0 : \kappa_{-j} = \kappa_j$, and with $K = diag(\kappa_{-h}, \ldots, \kappa_h)$, the estimate of β can be written as $\hat{\beta} = (X'KX)^1 X'Ky$. With $e_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}'$, the estimate of the trend is:

$$\hat{m}_t = e_1\hat{\beta} = w'y = \sum_{j=-h}^{h} w_j y_{t-j} \text{ with } w = KX(X'KX)^{-1}e_1$$

To conclude, the estimate of the trend $\hat{m}_t$ can be obtained applying the symmetric filter w to y_t[9]. Moreover, $X'w = e_1$ so:

$$\sum_{j=-h}^{h} w_j = 1, \quad \forall r \in [\![1,d]\!] : \sum_{j=-h}^{h} j^r w_j = 0$$

---

[9]w is symmetric due to the symmetry of the kernel weights κ_j.

Hence, the filter *w* preserve deterministic polynomial of order *d*.

5.1 Different kernels

In signal extraction, observations are generally weighted according to their distance from time t: this is the role of the kernel function. In the discrete case, a kernel function is a set of weights κ_j, $j = 0, \pm 1, \ldots, \pm h$ with $\kappa_j \geq 0$ and $\kappa_j = \kappa_{-j}$. An important class of kernels is the Beta kernels. In the discrete case, up to a proportional factor (so that $\sum_{j=-h}^{h} \kappa_j = 1$):

$$\kappa_j = \left(1 - \left| \frac{j}{h+1} \right|^r \right)^s$$

with $r > 0$, $s \geq 0$. It encompasses all kernels used in this report, except Henderson, trapezoidal and gaussian kernel. The following kernels are considered in this report:

- $r = 1, s = 0$ uniform kernel:

$$\kappa_j^U = 1$$

- $r = s = 1$ triangle kernel:

$$\kappa_j^T = \left(1 - \left| \frac{j}{h+1} \right| \right)$$

- $r = 2, s = 1$ Epanechnikov (or Parabolic) kernel:

$$\kappa_j^E = \left(1 - \left| \frac{j}{h+1} \right|^2 \right)$$

- $r = s = 2$ biweight kernel:

$$\kappa_j^{BW} = \left(1 - \left| \frac{j}{h+1} \right|^2 \right)^2$$

- $r = 2, s = 3$ triweight kernel:

$$\kappa_j^{TW} = \left(1 - \left| \frac{j}{h+1} \right|^2 \right)^3$$

- $r = s = 3$ tricube kernel:

$$\kappa_j^{TC} = \left(1 - \left| \frac{j}{h+1} \right|^3 \right)^3$$

- Henderson kernel (see section 5.1.1 for more details):

$$\kappa_j = \left[1 - \frac{j^2}{(h+1)^2} \right] \left[1 - \frac{j^2}{(h+2)^2} \right] \left[1 - \frac{j^2}{(h+3)^2} \right]$$

- Trapezoidal kernel:

$$\kappa_j^{TP} = \begin{cases} \frac{1}{3(2h-1)} & \text{if } j = \pm h \\ \frac{2}{3(2h-1)} & \text{if } j = \pm(h-1) \\ \frac{1}{2h-1} & \text{otherwise} \end{cases}$$

- Gaussian kernel[10]

$$\kappa_j^G = \exp \left(-\frac{j^2}{2\sigma^2 h^2} \right)$$

Henderson, trapezoidal and gaussian kernel are very specific:

---

[10]In this report we arbirarily take $\sigma^2 = 0.25$.

- The Henderson and trapezoidal kernel functions change with the bandwidth (the other kernel only depend on the ratio $j/h+1$).

- Other definitions of the trapezoidal and gaussian kernel can be used. The trapezoidal kernel is here considered because it corresponds to the filter used to extract the seasonal component in the X-12ARIMA algorithm. Therefore it is never used to extract trend-cycle component.

Figure 2: Coefficients of the different kernels for h from 2 to 30.
Note: to see the animation, the PDF must be open with Acrobat Reader, KDE Okular, PDF-XChange or Foxit Reader. Otherwise you will only be able to see the results for $h=2$.

The figure 2 summarizes the coefficients of the different kernels. Analyzing the coefficients we can already anticipate some properties of the associated filters:

- The triweight kernel has the narrowest distribution. The narrowest a distribution is, the smallest the weights of furthest neighbors are: the associated filter should have a high weight in the current observation (t).

- For h high the Henderson kernel is equivalent to the triweight kernel (since $h+1 \sim h+2 \sim h+3$, $\kappa_j^H \sim \kappa_j^{TW}$), the associated filter should also be equivalent. However, for h small ($h \leq 10$) the Henderson kernel is closer to the biweight kernel than to the triweight kernel.

5.1.1   Specific symmetric filters

When $p = 0$ (local adjustment by a constant) we obtain the **Nadaraya-Watson**'s estimator.

With the uniform kernel we obtain the **Macaulay filter**. When $p = 0, 1$, this is the arithmetic moving average: $w_j = w = \frac{1}{2h+1}$.

The **Epanechnikov** kernel is often recommended as the optimal kernel that minimize the mean square error of the estimation by local polynomial.

Loess is a locally weighted polynomial regression that use tricube kernel.

The **Henderson filter** is a specific case of a local cubic fit ($p = 3$), widely used for trend estimation (for example it's the filter used in the seasonal adjustment software X-12ARIMA). For a fixed bandwidth, Henderson found the kernel that gave the smoothest estimates of the trend. He showed that the three following problems were equivalent:

1. minimize the variance of third difference of the series by the application of the moving average;

2. minimize the sum of squares of third difference of the coefficients of the filter, it's the *smoothness criterion*: $S = \sum_j (\nabla^3 \theta_j)^2$;

3. fit a local cubic polynomial by weighted least squares, where the weights are chose to minimize the sum of squares of the resulting filter.

Resolving the last problem leads to the kernel presented in section 5.1.

5.1.2   Analysis of symmetric filters

In this section, all the filters are computed by local polynomial of degree $d = 3$. The figure 3 plots the coefficients of the filters for the different kernels presented in different kernels presented in section 5.1 and for different bandwidths h. The table 2 shows the variance reduction of the different filters. We find the similar results than in section 2:

- The triweight kernel gives the filter with the narrowest distribution. The narrowest a distribution is, the higher the variance reduction should be. Indeed, the distribution of the

coefficients of the filter can be interpreted as the output signal of an additive outlier. As a result, with a wide distribution, an additive outlier will be more persistent than with a narrow distribution. Therefore, it's the triweight that has the higher variance reduction for all $h \leq 30$.

- For h small, the trapezoidal filter seems to produce similar results than the Epanechnikov one.

- For h small the Henderson filter is closed to the biweight kernel, for h high it is equivalent to the triweight kernel.

Table 2: Variance reduction ratio ($\sum \theta_i^2$) of symmetric filters computed by local polynomial of degree 3.

					Kernel				
h	Biweight	Epanechnikov	Gaussian	Henderson	Trapezoidal	Triangular	Tricube	Triweight	Uniform
2	0.50	0.49	0.49	0.50	0.51	0.51	0.49	0.52	0.49
3	0.33	0.30	0.30	0.32	0.31	0.33	0.32	0.37	0.28
4	0.25	0.23	0.24	0.25	0.23	0.25	0.25	0.28	0.22
5	0.22	0.21	0.21	0.22	0.20	0.22	0.22	0.24	0.20
6	0.20	0.19	0.19	0.20	0.19	0.20	0.20	0.21	0.19
7	0.19	0.18	0.18	0.19	0.18	0.19	0.19	0.20	0.18
8	0.18	0.17	0.18	0.18	0.17	0.18	0.18	0.19	0.17
9	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.18	0.17
10	0.17	0.16	0.17	0.17	0.16	0.17	0.17	0.17	0.16
20	0.12	0.12	0.13	0.13	0.12	0.12	0.13	0.13	0.12
30	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10

Moreover, we find that for all the filters, the coefficients decrease, when the distance to the central observation increases, until a negative value and then increase towards 0 (except for the uniform kernel). Negative coefficients might be disturbing but they arise from the cubic polynomial constraints. Indeed to preserve polynomial of degree 2 (and so 3) we need $\sum_{j=-h}^{h} j^2 \theta_i = 0$, which constraint some coefficients to be negative. However, those negative coefficients are negligible compare to the central coefficients (they are more 80% smaller than the central coefficient for all kernels, except for uniform and trapezoidal with high bandwidth).

5.1.3 Gain functions

Figure 4 plots the gain functions of the different filters. Gain functions are usually plotted between 0 and π. However, locally weighted polynomial regression are low-pass filters: they leave almost unchanged low frequency components (such as the trend) and attenuate high frequency fluctuations (noise). For a monthly data, a cycle of 3 years corresponds to the frequency $2\pi/36$ and a cycle of 7 years to the frequency $2\pi/84$. Therefore, an ideal pass-band filter will have a gain function equal to 1 for low frequency ($\leq 2\pi/36$) and equal to 0 for other frequencies.

Figure 3: Coefficients of symmetric filters computed by local polynomial of degree 3, according to the differents kernels and for *h* from 2 to 30.
Note: to see the animation, the PDF must be open with Acrobat Reader, KDE Okular, PDF-XChange or Foxit Reader. Otherwise you will only be able to see the results for h = 2.

When the bandwidth *h* increases, the gain function decreases for low frequencies: short business cycles will then be attenuated. For a fixed value of *h*, gaussian, Henderson and triweight filters will preserve more short business cycles than the other filters (especially uniform, trapezoidal and Epanechnikov). Moreover, the gain function of those filters decreases faster to zero with less fluctuations: it enhances the higher variance reduction ratio shown in table 2.

Just analyzing the symmetric filters properties, there is no doubt that Henderson, triweight and biweight filters have similar properties and will perform better than the other kernel for

Figure 4: Gain functions from 0 to $2\pi/12$ of symmetric filters computed by local polynomial of degree 3, according to the differents kernels and for *h* from 2 to 30.
Note: the two horizontal lines corresponds to the frequencies $2\pi/84$ (cycle of 7 years) and $2\pi/36$ (cycle of 3 years). to see the animation, the PDF must be open with Acrobat Reader, KDE Okular, PDF-XChange or Foxit Reader. Otherwise you will only be able to see the results for $h = 2$.

trend-cycle extraction. The same results are found with asymmetric filters. Thus, in order to simplify the presentation analysis, in the next sections we will only show the results with the Henderson filter.

5.2 Asymmetric filters

5.2.1 Direct asymmetric filters (DAF)

As mentioned in section 4.3, symmetric filters cannot be used in boundary points. For real-time estimation, three different approaches can be used:

1. Build an asymmetric filter fitting local polynomial to the available observations y_t for $t \in [\![n-h,n]\!]$.

2. Apply the symmetric filter to the series extended by forecast (or backcast) $\hat{y}_{n+l|n}, l \in [\![1,h]\!]$.

3. Build an asymmetric filter which minimize the mean square revision error subject to polynomial reproducing constraints.

Proietti and Luati (2008) show that the first two approaches are equivalent when the forecast is done by a polynomial extrapolation of order d (forecasts generated with the same polynomial model than the symmetric filter). This is called the *direct asymmetric filter* (DAF). Let q be the number of available observations in the future: q varies from 0 (real time filter) to h (symmetric filter).

Rewriting the matrix X, K y in the following way:

$$X = \begin{pmatrix} X_p \\ X_f \end{pmatrix}, \quad y = \begin{pmatrix} y_p \\ y_f \end{pmatrix}, \quad K = \begin{pmatrix} K_p & 0 \\ 0 & K_f \end{pmatrix}$$

where y_p correspond to the available data and y_f the missing data. The DAF w_a and the forecast $\hat{y}_f$ can be written as:

$$w_a = K_p X_p (X_p' K_p X_p)^{-1} e_1, \quad \hat{y}_f = X_f (X_p' K_p X_p)^{-1} X_p' K_p y_p$$

Moreover, we have the following results with the DAF w_a:

- it satisfies the same polynomial reproduction constraints as the symmetric filter (conserve polynomial of degree d). Thus, the bias in estimating an unknown function of time has the same order of magnitude as in the interior of time support.

- w_a minimize the weighted distance (by the kernel function) between the asymmetric filter coefficients and the symmetric ones. Therefore, for the DAF it is equivalent to fit a local polynomial and to minimize the revisions

However, the weights $w_{a,0}$ of the DAF are highly concentrated in the current observation t with an important change between $q = 0$ (real-time filter) and $q = h$ (see figure 5). Moreover the real-time filter doesn't have a satisfying gain functions: it is closer to one for all the frequencies (it thus has a low noise reduction power). Therefore, even if the real-time filter is unbiased (if the series is generated by a polynomial of degree d) it is at the expenses of a high variance.
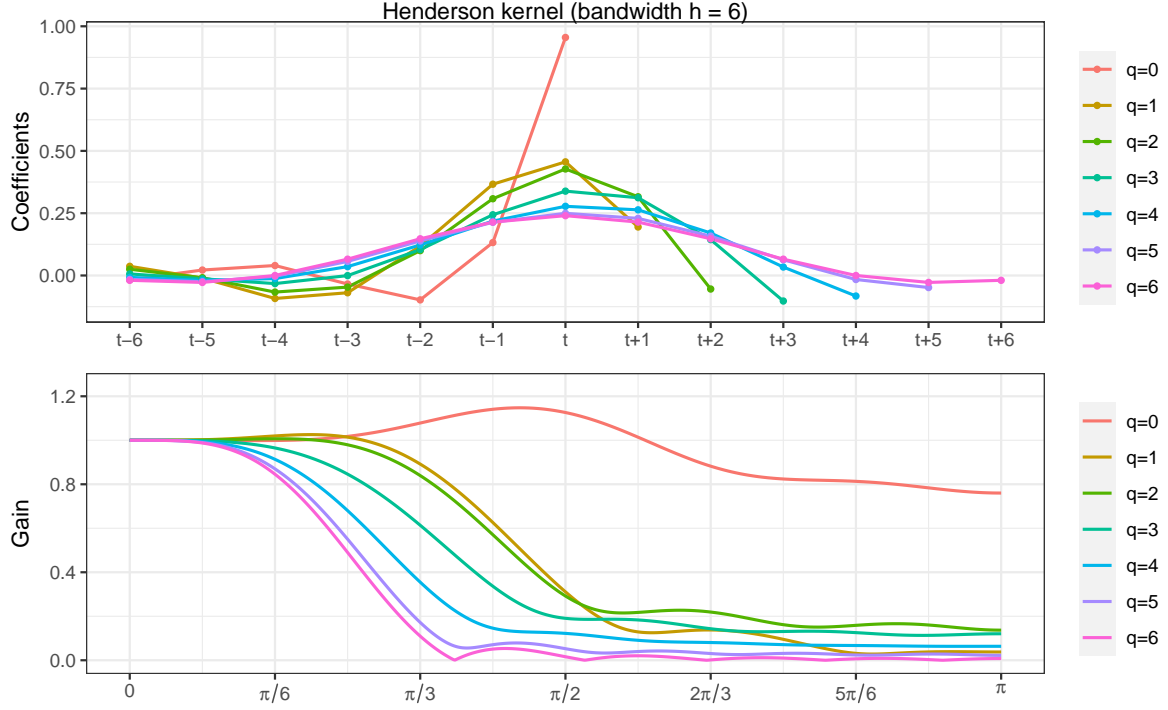
15

Figure 5: Coefficients and gain function of direct asymmetric filters (DAF) computed by local polynomial of degree 3 with the Henderson kernel for $h = 6$.

For all the kernels, we find the same results as in Proietti and Luati (2008):

- For a fixed value of d, the more the data is available (q increases), the more the weight associated to the current observation $w_{a,0}$ decreases.

- For a fixed value of h and q, $w_{a,0}$ increases exponentially with the polynomial degree d (in particular, for $d = h$, $w_{a,0} = 1$).

5.2.2 General class of asymmetric filters

To deal with the problem of the variance of the estimates of the real-time filters, Proietti and Luati (2008) suggest a general of asymmetric filters to make a tradeoff between bias and variance.

Here we consider that the data is generated by the model:

$$y = U\gamma + Z\delta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, D)$$

The goal is to find a filter v which minimize the mean square revision error (with the symmetric filter w) subject to some constraints. The constraints are summarized by the matrix $U = \begin{pmatrix} U'_p & U'_f \end{pmatrix}'$ (with U_p the available observations of the matrix U for the asymmetric filter):

$U_p'v = U'w$. The problem is equivalent to find v that minimize:

$$\varphi(v) = \underbrace{\underbrace{(v-w_p)'D_p(v-w_p) + w_f'D_f w_f}_{\text{revision error variance}} + \underbrace{[\delta'(Z_p'v - Z'w)]^2}_{biais^2}}_{\text{Mean square revision error}} + \underbrace{2l'(U_p'v - U'w)}_{\text{constraints}} \qquad (1)$$

with l a vector of Lagrange multipliers.

When $U = X$ this is equivalent to the constraint to preserve polynomial of degree d: we find the direct asymmetric filters w_a with $D = K^{-1}$.

When $U = \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix}'$, $Z = \begin{pmatrix} -h & \cdots & +h \end{pmatrix}'$, $\delta = \delta_1$, $D = \sigma^2 I$ and when the symmetric filter is the Henderson filter we obtain the Musgrave asymmetric filters (used in the seasonal adjustment algorithm X-12ARIMA, see Musgrave (1964)). With this filter we assume that the data is generated by a linear process and that the asymmetric filters preserve constant signals ($\sum v_i = \sum w_i = 1$). The asymmetric filters depends on the ratio δ_1/σ, which is related to the "I-C" ratio $R = \frac{\bar{I}}{\bar{C}} = \frac{\sum|I_t - I_{t-1}|}{\sum|C_t - C_{t-1}|}$ ($\delta_1/\sigma = 2/(R\sqrt{\pi})$), the ratio between the expected absolute difference of the irregular and of the trend-cycle. In the seasonal adjustment method, the I-C ratio[11] is used to determine the bandwidth to use for the Henderson filter. For monthly data:

- if $R < 1$ a 9-term Henderson is used ($h = 4$);

- if $1 \le R \le 3.5$ a 13-term Henderson is used ($h = 6$);

- if $3.5 < R$ a 23-term Henderson is used ($h = 12$).

In this report, for simplicity we only consider 13-term symmetric filters: the ratio δ^2/σ^2 is fixed to 3.5.

When U corresponds to the first $d^* + 1$ first the columns of X, $d^* < d$, the constraint is that the asymmetric filter should reproduce polynomial of degree d^*, the potential bias depends on the value of δ. This will reduce the variance at the expense of a bias: it is the idea followed by Proietti and Luati (2008) to propose three class of asymmetric filters:

1. *Linear-Constant* (LC): y_t linear ($d = 1$) and v preserve constant signals ($d^* = 0$). We obtain Musgrave filters when the Henderson kernel is used.

2. *Quadratic-Linear* (QL): y_t quadratic ($d = 2$), v depend on the ratio δ_2^2/σ^2.

3. *Cubic-Quadratic* (CQ): y_t cubic ($d = 3$), v depend on the ratio δ_2^2/σ^2.

The table 3 shows the quality criteria of the different methods with the Henderson kernel and $h = 6$. For real-time filters ($q = 0$), the more complex the filter is (in terms of polynomial preservation), the less the timeliness is and the more the fidelity/smoothness is: the reduction

---

[11]To compute the I-C ratio, a first decomposition of the seasonally adjusted series is computed using a 13-term Henderson moving average.

of the time-delay is at the expense of an increase variance. This change when q increases: for $q = 2$ the QL filter has a greater timeliness that the LC filter. This unexpected result underlines the fact that in the approach of Proietti and Luati (2008), the timeliness is never set as a goal to minimize.

Table 3: Quality criteria of asymmetric filters ($q = 0, 1, 2$) computed by local polynomial with Henderson kernel for $h = 6$ and $R = 3.5$.

Method	b_c	b_l	b_q	F_g	S_g	$T_g \times 10^{-3}$	A_w	S_w	T_w	R_w
$q = 0$										
LC	0	-0.407	-2.161	0.388	1.272	30.341	0.098	0.488	0.409	0.548
QL	0	0.000	-0.473	0.711	5.149	0.047	0.067	1.894	0.000	0.106
CQ	0	0.000	0.000	0.913	11.942	0.015	0.016	2.231	0.000	0.102
DAF	0	0.000	0.000	0.943	14.203	0.003	0.015	2.178	0.000	0.098
$q = 1$										
LC	0	-0.121	-0.525	0.268	0.433	4.797	0.009	0.119	0.063	0.112
QL	0	0.000	-0.061	0.287	0.707	0.694	0.005	0.192	0.007	0.042
CQ	0	0.000	0.000	0.372	0.571	0.158	0.022	0.575	0.001	0.061
DAF	0	0.000	0.000	0.409	0.366	0.061	0.020	0.760	0.000	0.059
$q = 2$										
LC	0	0.003	1.076	0.201	0.080	0.347	0.009	0.012	0.004	0.015
QL	0	0.000	0.033	0.215	0.052	2.083	0.000	0.011	0.023	0.067
CQ	0	0.000	0.000	0.370	0.658	0.131	0.021	0.558	0.001	0.055
DAF	0	0.000	0.000	0.398	0.768	0.023	0.017	0.677	0.000	0.048

Regarding the mean square revision error ($A_w + S_w + T_w + R_w$), LC and QL filters always gives better results than CQ and DAF filters. This "theorical" mean square revision error can be compared to the "empirical" one computed applying the filters to real data. The table 4 shows the average mean square revision error of the different filters applied to the Industrial production indices of the European Union between 2003 and 2019. The mean square revision errors are different in level but give same results comparing the different filters: this validate the formula used for the criteria $A_w, S_w, T_w,$ and R_w (see section 7).

Taking for the value R the "I-C" ratio computed with the raw data gives the same summary results.

➲ All those results suggest focusing on LC and QL filters and to focus on asymmetric linear filters that preserve polynomial trends of degree less than one.

The results for the different kernels can also be visualized in an online application available at https://aqlt.shinyapps.io/FiltersProperties/.

Table 4: Mean square revision error of asymmetric filters ($q = 0, 1, 2$) computed by local polynomial on the Industrial production indices of the European Union.

Method	Mean squared revision error		$A_w + S_w + T_w + R_w$
	2007-2010	2003-2019	
$q = 0$			
LC	1.54	0.855	1.543
QL	2.04	1.560	2.068
CQ	2.88	2.363	2.349
DAF	3.03	2.489	2.290
$q = 1$			
LC	2.96	1.315	0.304
QL	2.31	1.074	0.247
CQ	2.62	1.342	0.660
DAF	2.72	1.437	0.840
$q = 2$			
LC	6.78	2.774	0.040
QL	7.45	3.074	0.101
CQ	7.99	3.620	0.635
DAF	7.94	3.618	0.742

Note: the filters are computed with $h = 6$ (13-terms symmetric filter) and $R = 3.5$.

Summary — Local polynomial filters

Advantages:

- Simple models with an easy interpretation.

- The asymmetric linear filter is independent of the data and of the date of estimation.

Drawbacks:

- Timeliness is not controlled.

---

6 General optimization problem: FST filters

6.1 Description of the approach

Grun-Rehomme, Guggemos, and Ladiray (2018) defined a general approach to derive linear filters, based on an optimization problem of three criteria:

- *Fidelity*, F_g: it's the variance reduction ratio. It is called "Fidelity" because we want the output signal to be as close as possible to the input signal where the noise component is

removed

$$F_g(\theta) = \sum_{k=-p}^{+f} \theta_k^2$$

F_g can be rewritten in a positive quadratic form: $F_g(\theta) = \theta'F\theta$ with F the identity matrix of order $p + f + 1$.

- *Smoothness*, S_g: it's the Henderson smoothness criterion (sum of the squared of the third difference of the coefficients of the filter). It measures the flexibility of the coefficient curve of a filter and the smoothness of the trend.

$$S_g(\theta) = \sum_j (\nabla^3 \theta_j)^2$$

S_g could also be rewritten in a positive quadratic form: $S_g(\theta) = \theta'S\theta$ with S a symmetric matrix of order $p + f + 1$.

- *Timeliness*, T_g: it measures the phase shift between input and output signal for specific frequencies. When a linear filter is applied, the level input signal is also altered by the gain function. Therefore, it is natural to consider that the higher the gain is, the higher the phase shift impact is. That's why the timeliness criterion depends on the gain and phase shift functions (ρ_θ and φ_θ), the link between both functions being made by a penalty function f.

$$T_g(\theta) = \int_{\omega_1}^{\omega_2} f(\rho_\theta(\omega), \varphi_\theta(\omega)) \, d\omega$$

In this article we use $\omega_1 = 0$ and $\omega_2 = 2\pi/12$ (for monthly data): we focus on the phase shift impact on cycles of more than one year. For the penalty function, we take $f: (\rho, \varphi) \mapsto \rho^2 \sin(\varphi)^2$. Indeed, for this function, the timeliness criterion is analytically solvable ($T_g = \theta'T\theta$ with T a square symmetric matrix of order $p + f + 1$), which is better in a computational point of view.

The asymmetric filters are computed minimizing a weighted sum of the past three criteria, subject to some constraints. Those constraints are usually polynomial preservation.

$$\begin{cases} \min_\theta & J(\theta) = \alpha F_g(\theta) + \beta S_g(\theta) + \gamma T_g(\theta) \\ s.t. & C\theta = a \end{cases}$$

The conditions $\alpha, \beta, \gamma \geq 0$ and $\alpha + \beta \neq 0$ guarantee that $J(\theta)$ is a strictly convex function: therefore the optimization problem has a unique solution.

The Henderson symmetric filters can for example be computed with

$$C = \begin{pmatrix} 1 & \cdots & 1 \\ -h & \cdots & h \\ (-h)^2 & \cdots & h^2 \end{pmatrix}, \quad a = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \alpha = \gamma = 0, \quad \beta = 1$$

This approach can be called as the "FST approach" in reference to the three indicators used in the optimization problem.

Summary — FST filters

Advantages:

- The asymmetric linear filter is independent of the symmetric filter, the data and the date of estimation.

- Unique solution to the optimization problem.

- The approach can be customized adding new criteria.

Drawbacks:

- The different criteria are not normalized: the associated weights cannot be compared.

6.2 Extension with the revision criterion

The FST — Fidelity-Smoothness-Timeliness — approach is the only one that doesn't directly include a criterion on the revision error relative to a symmetric filter.

This approach could be somewhat generalized in order to include the revision criterion, replacing in the orthogonal form θ by $(w - \theta)$ (with w the symmetric filter), and F by a matrix relative to the data (see Ladiray (2018)).

In this paper, we consider the method that consists to extend the minimization problem of local polynomial filters adding the Timeliness criterion defined by Grun-Rehomme, Guggemos, and Ladiray (2018)[12]. Using the same notations as in section 5.2.2, $\theta = v$ and noting $g = v - w_p$, the Timeliness criterion can be rewritten:

$$T_g(v) = v'Tv = g'Tg + 2w_p'Tg + w_p'Tw_p \quad (T \text{ being symmetric})$$

---

[12]This method is for example coded in Java by Jean Palate in https://github.com/palatej/jdemetra-core.

Moreover, the objective function φ of equation (1) can be rewritten as:

$$\varphi(v) = (v - w_p)' D_p (v - w_p) + w_f' D_f w_f + [\delta'(Z_p' v - Z' w)]^2 + 2l'(U_p' v - U' w)$$

$$= g'Qg - 2Pg + 2l'(U_p' v - U' w) + c \quad \text{with} \begin{cases} Q = D_p + Z_p \delta \delta' Z_p' \\ P = w_f Z_f \delta \delta' Z_p' \\ c \text{ a constant independent of } v \end{cases}$$

Adding the Timeliness criterion, it becomes:

$$\widetilde{\varphi}(v) = g'\widetilde{Q}g - 2\widetilde{P}g + 2l'(U_p' v - U' w) + \widetilde{c} \quad \text{with} \begin{cases} \widetilde{Q} = D_p + Z_p \delta \delta' Z_p' + \alpha_T T \\ \widetilde{P} = w_f Z_f \delta \delta' Z_p' - \alpha_T w_p T \\ \widetilde{c} \text{ a constant independent of } v \end{cases}$$

where α_T is the weight associated to the Timeliness criterion. With $\alpha_T = 0$ we find $\varphi(v)$.

The figures 6 show the impact of α_T on the coefficients of the linear filter with the LC method:

- The more α_T increases, the more the coefficient associated to the current observation increases: this is what we expected.

- α_T impacts logarithmically the coefficients: we can restraint α_T to $[0, 2000]$.

- As expected, including the timeliness criterion has more impact for the value of q that gives filters with higher timeliness: it corresponds to $q \leq 2$ for the LC method. For the QL method we find that α_T has an impact for medium values of q ($2 \leq q \leq 4$).

7 Data-dependent filter

In Wildi and McElroy (2019), the authors proposed a data-dependent approach to derive linear filters. They decompose the mean square revision error in a trilemma between three quantities: *accuracy*, *timeliness* and *smoothness*.

Let:

- $\{x_t\}$ be our input time series;

- $\{y_t\}$ the target signal, i.e. the result of a symmetric filter, and Γ_s, ρ_s and φ_s the associated frequency response, gain and phase shift functions.

- $\{\hat{y}_t\}$ an estimation of $\{y_t\}$, i.e. the result of an asymmetric filter (when not all observations are available), and Γ_θ, ρ_θ and φ_θ the associated frequency response, gain and phase shift functions.

Figure 6: Impact of the timeliness weight (α_T) on the coefficients of the local polynomial filter with the LC method with $h = 6$, $R = 3.5$ and the Henderson kernel.
Note: to see the animation, the PDF must be open with Acrobat Reader, KDE Okular, PDF-XChange or Foxit Reader. Otherwise you will only be able to see the results for $q = 0$.

If we assume that $\{x_t\}$ is weakly stationary with a continuous spectral density h, the mean square revision error, $\mathbb{E}\left[(y_t - \hat{y}_t)^2\right]$, can be written as:

$$\mathbb{E}\left[(y_t - \hat{y}_t)^2\right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 h(\omega) \, d\omega = \frac{1}{2\pi} \times 2 \times \int_0^{\pi} |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 h(\omega) \, d\omega$$

(2)

This equality can also be generalized to non-stationary integrated process (for example imposing cointegration between both signals and using pseudo-spectral density, see Wildi and Mcelroy (2013)).

We have:

$$|\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 = \rho_s(\omega)^2 + \rho_\theta(\omega)^2 + 2\rho_s(\lambda)\rho_\theta(\lambda)\left(1 - \cos(\varphi_s(\omega) - \varphi_\theta(\omega))\right)$$

$$= (\rho_s(\omega) - \rho_\theta(\omega))^2 + 4\rho_s(\lambda)\rho_\theta(\lambda)\sin^2\left(\frac{\varphi_s(\omega) - \varphi_\theta(\omega)}{2}\right)$$

(3)

23

The interval $[0, \pi]$ is then splitted in two: the pass-band $[0, \omega_1]$ (the frequency interval that contains the target signal) and the stop-band $[\omega_1, \pi]$.

The mean squared error defined in equation (2) can then be decomposed additively into four quantities:

$$Accuracy = A_w = 2 \int_0^{\omega_1} (\rho_s(\omega) - \rho_\theta(\omega))^2 h(\omega) \, d\omega$$

$$Timeliness = T_w = 8 \int_0^{\omega_1} \rho_s(\lambda) \rho_\theta(\lambda) \sin^2 \left(\frac{\varphi_\theta(\omega)}{2} \right) h(\omega) \, d\omega$$

$$Smoothness = S_w = 2 \int_{\omega_1}^{\pi} \left(\rho_s(\omega)^2 - \rho_\theta(\omega) \right)^2 h(\omega) \, d\omega$$

$$Residual = R_w = 8 \int_{\omega_1}^{\pi} \rho_s(\lambda) \rho_\theta(\lambda) \sin^2 \left(\frac{\varphi_\theta(\omega)}{2} \right) h(\omega) \, d\omega$$

Remark. *To have coherent definitions between all sections, the formulas of the four criteria slightly differ from the ones defined in Wildi and McElroy (2019):*

- *in this paper the interval integrations is $[0, \pi]$ rather than $[-\pi; \pi]$ (the integrals are then only multiplied by 2 because all the functions are even);*

- *the pass-pand interval is defined as the frequency interval that contains the target signals whereas in Wildi and McElroy (2019) it depends on the gain function of the symmetric filer (pass-band$= \{\omega | \rho_s(\omega) \geq 0.5\}$).*

In general, the residual R_w is small since $\rho_s(\omega) \rho_\theta(\omega)$ is close to 0 in the stop-band. Moreover, user priorities are rarely concerned about the time-shift properties of components in the stop-band. That's why, to derive linear filters the residual is not taken into account; and that's why the authors suggest minimizing a weighted sum of the first three indicators:

$$\mathscr{M}(\vartheta_1, \vartheta_2) = \vartheta_1 T_w(\theta) + \vartheta_2 S_w(\theta) + (1 - \vartheta_1 - \vartheta_2) A_w(\theta)$$

One of the drawbacks of this method is that there is no guarantee that there is a unique solution.

In this paper we focus in non-parametric approaches to derive linear filters. That's why this approach is not considered. However, the decomposition of the mean squared error gave useful indicators to compare linear filters because, in contrary the one presented in section 6, their values can be easily interpreted and compared to each other.

To have criteria that don't depend on the data (and the one defined in table 1), we take for the symmetric filter the Henderson filter and we fix the spectral density to the one of a random walk:

$$h_{RW}(x) = \frac{1}{2(1 - \cos(x))}$$

24

> **Summary — Data-dependent filters**
>
> **Advantages**:
>
> - The values of the different criteria can be compared: the weights can be easily interpreted.
>
> **Drawbacks**:
>
> - Data-dependent filter: it depends on the symmetric filter, the data and the date of estimation.
>
> - Some optimization problems might occur (several minimum, etc.).

We could also use a non-parametric approach to build asymmetric using h_{RW} for the spectral density of the series. However, due to a lack of time, this option is not considered in this study.

8 Asymmetric filters and Reproducing Kernel Hilbert Space

Dagum and Bianconcini ([2008](#)) developed a general approach to derive linear filters based on Reproducing Kernel Hilbert Space (RKHS) methodology.

A RKHS is a Hilbert space characterized by a kernel that reproduces, via an inner product defined by a density function $f_0(t)$, every function of the space. Therefore, a kernel estimator K_p of order p (i.e.: that reproduce without distortion a polynomial trend of degree $p-1$) can be decomposed into the product of a reproducing kernel R_{p-1}, belonging to the space of polynomials of degree $p-1$, and a probability density function f_0.

For any sequence $(P_i)_{0 \leq i \leq p-1}$ of orthonormal polynomials in $\mathbb{L}^2(f_0)$[13], the kernel estimator K_p is defined by:

$$K_p(t) = \sum_{i=0}^{p-1} P_i(t) P_i(0) f_0(t)$$

The weights of a symmetric filter are then derived with:

$$\forall j \in [\![-h, h]\!] \ : \ w_j = \frac{K_p(j/b)}{\sum_{i=-h}^{h} K_p(i/b)}$$

where b is a time-invariant global bandwidth parameter.

---

[13]$\mathbb{L}^2(f_0)$ is the Hilbert space defined by the inner product:

$$\langle U(t), V(t) \rangle = \mathbb{E}\left[U(t)V(t)\right] = \int_{\mathbb{R}} U(t)V(t) f_0(t) \, \mathrm{d}t$$

The density f_0 corresponds to the continuous versions of the kernel defined in 5.1. For example, the biweight function is $f_{0B}(t) = (15/16)(1-t^2)^2, t \in [-1,1]$. The local polynomial filter obtained with the biweight kernel is then obtained using the bandwidth $b = m+1$.

The goal of Dagum and Bianconcini (2008) is to derive asymmetric filters from the Henderson symmetric filter, therefore the ideal kernel function would be the Henderson one. However, as shown in section 5.1, the Henderson density is a function of the bandwidth and needs to be calculated any time m changes (as its corresponding orthonormal polynomial). That's why the authors use the biweight kernel to approximate the Henderson kernel (for $h \geq 24$ they suggest to consider the triweight kernel).

The asymmetric weighted are obtained adapting the kernels to the length of the asymmetric filters:

$$\forall j \in [\![-h, p]\!] : w_{a,j} = \frac{K_p(j/b)}{\sum_{i=-h}^{p} K_p(i/b)}$$

With $b = h+1$, Proietti and Luati (2008) show that we obtain the direct asymmetric filters (DAF).

In Bee Dagum and Bianconcini (2015), the authors suggest performing an optimal bandwidth selection (parameter b), decomposing the mean squared revision error as in equation (3) but with a uniform spectarl density ($h(\omega) = 1$). The bandwidth can then be chosen to minimize the mean squared revision error, the phase shift, etc. The following bandwidth selection are studied:

$$b_{q,G} = \min_{b_q \in]h;2h+1]} \sqrt{2 \int_0^\pi (\rho_s(\omega) - \rho_\theta(\omega))^2 \, d\omega}$$

$$b_{q,\gamma} = \min_{b_q \in]h;2h+1]} \sqrt{2 \int_0^\pi |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 \, d\omega}$$

$$b_{q,\varphi} = \min_{b_q \in]h;2h+1]} 8 \int_0^{2\pi/12} \rho_s(\lambda)\rho_\theta(\lambda) \sin^2 \left(\frac{\varphi_\theta(\omega)}{2} \right) d\omega$$

One of the drawbacks of the optimal bandwidth selection is that there is no guarantee that there is a unique solution.

Remark. *The formulas of $b_{q,G}$, $b_{q,\gamma}$ and $b_{q,\varphi}$ slightly differ from the ones defined in Bee Dagum and Bianconcini (2015) to have coherent definitions between all sections:*

- *$b_{q,\varphi}$ is defined as*

$$b_{q,\varphi} = \min_{b_q \in]h;2h+1]} \sqrt{2 \int_{\Omega_S} \rho_s(\lambda)\rho_\theta(\lambda) \sin^2 \left(\frac{\varphi_\theta(\omega)}{2} \right) d\omega}$$

With $\Omega_S = [0, 2\pi/36]$ the frequency domain associated to cycles of 16 months or longer.

- *in Bee Dagum and Bianconcini (2015) a different formula is used for the frequency*

response function ($\Gamma_\theta(\omega) = \sum_{k=-p}^{+f} \theta_k e^{2\pi i \omega k}$): it only changes the interval integration.

The table 5 shows the quality criteria of the RKHS filters. Even if the symmetric filter preserves quadratic trends, this is not the case of the asymmetric filters: they only preserve constant trends. The more data is available (i.e.: q increases), the less the asymmetric filters distort polynomial trends (for example, for $q = 2$, $b_c \simeq 0$).

Table 5: Quality criteria of asymmetric filters ($q = 0, 1, 2$) computed by the RKHS methodology $h = 6$.

Method	b_c	b_l	b_q	F_g	S_g	$T_g \times 10^{-3}$	A_w	S_w	T_w	R_w
$q = 0$										
$b_{q,\gamma}$	0	-1.526	3.893	0.222	0.469	74.991	0.017	0.069	2.164	0.671
$b_{q,G}$	0	-2.039	6.937	0.177	0.305	98.650	0.082	0.066	3.547	0.715
$b_{q,\varphi}$	0	-0.614	0.033	0.381	1.204	23.323	0.019	0.549	0.451	0.356
$q = 1$										
$b_{q,\gamma}$	0	-0.516	0.992	0.226	0.303	14.386	0.002	0.047	0.313	0.170
$b_{q,G}$	0	-0.923	2.880	0.189	0.248	27.386	0.039	0.036	0.786	0.213
$b_{q,\varphi}$	0	-0.176	0.295	0.276	0.359	2.901	0.003	0.182	0.050	0.071
$q = 2$										
$b_{q,\gamma}$	0	0.041	0.863	0.202	0.090	0.328	0.006	0.010	0.004	0.018
$b_{q,G}$	0	-0.007	1.001	0.197	0.096	0.637	0.009	0.010	0.008	0.021
$b_{q,\varphi}$	0	0.105	0.820	0.215	0.077	0.064	0.003	0.025	0.005	0.010

Summary — RKHS filters

Advantages:

- Filters that apply to irregular frequency series (for example with a lot of missing values) can easily be computed.

Drawbacks:

- The linear filters don't preserve polynomial trends of degree 1 or more.

- Some optimization problems might occur (several minimum, etc.).

---

9 Comparison of the different filters

9.1 Comparison with the FST approach

The FST approach provides a useful tool to validate a method of construction of linear filter. Indeed, a linear filter can be considered as suboptimal if we can find a set of weight for the FST

approach that gives better results (in terms of fidelity, smoothness and timeliness) with the same (or higher) polynomial constraints.

For $h = 6$ (13-term symmetric filter), for the RKHS filters we find that:

- imposing that asymmetric filter preserve constants, the FST approach gives better results for all values of q for the filters computed with $b_{q,G}$ and $b_{q,\gamma}$, and for $q \leq 3$ for the one computed with $b_{q,\varphi}$.

- imposing that asymmetric filter preserve linear trends, the FST approach gives better results for $q \in [2,5]$ for the filters computed with $b_{q,G}$ and $b_{q,\gamma}$, and for $q \in [2,4]$ for the one computed with $b_{q,\varphi}$.

➲ Therefore, the RKHS approach seems not to be the better approach to build asymmetric filters to apply to monthly regular data (i.e.: data for which we usually use $h = 6$).

For the local polynomial filters we find that:

- imposing that asymmetric filter preserve constants, the FST approach gives better results than the LC method for all values of q. Adding the timeliness criterion allows to have better results than the FST approach for $q = 0, 1$.

- imposing that asymmetric filter preserve linear trends, the FST approach gives better results for $q \in [2,5]$ for the LC and QL methods. Adding the timeliness criterion doesn't change the results.

➲ In terms of quality criteria, local polynomial seems to perform better for real-time and one-period ahead filters ($q = 0, 1$).

Those comparisons can be seen in an online application available at https://aqlt.shinyapps.io/FSTfilters/.

9.2   Illustration with an example

In this section we compare the different asymmetric filters applying them to the French industrial production index of manufacture of motor vehicles, trailers and semi-trailers[14]. For simplicity, for the FST approach we only show an extreme filter example minimizing the phase shift by putting a large weight to the timeliness ($\alpha = 0$, $\beta = 1$, $\gamma = 1000$). For local polynomial filters, the Henderson kernel is used with $R = 3.5$. The figure 7 shows the results around a turning point (the economic crisis of 2009) and during "normal cycles" (between 2000 and 2008).

For this example, in terms of time lag to detect the turning point of February 2009[15], the DAF

---

[14]This series is produced by the INSEE, the French National Statistical Office, and is available at https://bdm.insee.fr/series/sdmx/data/SERIES_BDM/010537940.

[15]A turning point (upturn) is defined to occur at time t if $y_{t-k} \geq \cdots \geq y_{t-1} < y_t \leq y_{t+1} \leq \cdots y_{t+m}$. A downturn would be equivalently defined as a date t $y_{t-k} \leq \cdots \leq y_{t-1} > y_t \geq y_{t+1} \geq \cdots y_{t+m}$. Following Zellner, Hong, and

Figure 7: Application of asymmetric filters to the French industrial production index of manufacture of motor vehicles, trailers and semi-trailers.
Note: The final trend is the one computed by a symmetric 13-terms Henderson filter
To see the animation, the PDF must be open with Acrobat Reader, KDE Okular, PDF-XChange or Foxit Reader.
Otherwise you will only be able to see the results for the DAF filter.

is the worst filters since it introduces a delay of one month until $q = 3$. LC, QL, $b_{q,\gamma}$ filters introduce a delay only for real-time filter ($q = 0$) of one month except for the QL filter for which it is of two months. Surprisingly, the FST filter that minimize the phase shift introduce a time lag for $q = 0$ and $q = 1$ (the upturn is detected earlier), whereas $b_{q,\phi}$ doesn't introduce any time lag.

In terms of revision error, the DAF filters also give the worst results. For real-time estimates ($q = 0$), the LC and RKHS filters minimize the revision between 2000 and 2008 but when more data are available all the methods (except the DAF filters) give similar results. As excepted, the QL filter minimize the revision around the turning point of 2009 since it is the only filter that preserve polynomial trends of degree 1. The fact that, around the turning point, $b_{q,\phi}$ performs better than the other RKHS filters can be explain by the fact that this filter distort less polynomial trends of degree 1 and 2 (b_l and b_q statistics are lower, table 3).

Min (1991), $k = 3$ and $m = 1$ are chosen given the smoothness of the trend-cycle data.

With this simple example we cannot generalize properties on the different asymmetric filters, but it gives some leads to extend this study:

- minimize the timeliness doesn't imply earlier detection of turning points: more investigations on the timeliness criterion could be done;

- accept some bias in polynomial trends preservation can lead to asymmetric filters that performs better in term of turning points detection;

- for almost all the filters, there is no time lag to detect the turning points when two future observations are available, and the different methods gives similar results. Therefore, it is pointless to optimize asymmetric filters when too much future observations are available.

Moreover, this example also illustrate that the different methods performs better than the DAF filters. The real-time estimates methods that use this filter could be easily improved. This is for example the case of STL (Seasonal-Trend decomposition based on Loess) seasonal adjustment method proposed by Cleveland et al. (1990).

10   Conclusion

To sum up, several lessons can be learn from this study on asymmetric filters.

First, to build asymmetric filters there is no need to seek to preserve polynomial trends of degree more than one: it introduces more bias in the detection of turning point with a low gain on the revision error. Moreover, introduce a small bias in polynomial preservation can lead to filters that performs globally better, as we have shown with RKHS $b_{q,\phi}$ filter with the French IPI of manufacture of motor vehicles, trailers and semi-trailers.

Second, focusing on the Fidelity-Smoothness-Timeliness criteria, the FST approach can lead to asymmetric filters that outperform the other ones (local polynomials and RKHS). More research should be done in that direction to understand when it performs better and if the improvement is statistically significant.

Third, all the methods lots of degrees of freedom on the criteria used to build them. Therefore, more theoretical and empirical studies should be done on the effects of this degrees of freedom to see how they impact the estimation and if we cannot simplify them, as it is done in the X-12ARIMA algorithm to choose the length of the Henderson filter.

Finally, the study on asymmetric filters should be focus on real-time estimates and when few future observations are available. Indeed, for monthly data, when two or more future observations are available, the different methods give similar results and the gain to compute optimize filters is small. Consequently, adding the timeliness criterion to local polynomial filters will mainly be useful for the LC methods for which it has an impact when few future observations are known.

This study could also be extended in many ways. One lead could be to do more investigation on the length of the filters. In most of studies on asymmetric filters, the asymmetric filter uses as many past observations as the symmetric filters. However, we could consider asymmetric filters that takes more past observations. Moreover, in this paper we have only focus on regular monthly data, but we could have some issues with high frequency data and irregular data. In addition to the problem of choosing the length of the filter, we could check if we find similar results as for short filters.

Another lead could be to study the impact of atypical points: in X-12ARIMA there is a strong correction of atypical points on the irregular before applying the filters to extract the trend. This therefore prompts us to study what would be their impact on the estimation of trends and turning points, but also to explore new kinds of asymmetric filters based on robust methods.

References

Bee Dagum, Estela and Silvia Bianconcini (Sept. 2015). "A new set of asymmetric filters for tracking the short-term trend in real-time". In: *Ann. Appl. Stat.* 9.3, pp. 1433–1458. DOI: 10.1214/15-AOAS856. URL: https://doi.org/10.1214/15-AOAS856.

Cleveland, Robert B. et al. (1990). "STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion)". In: *Journal of Official Statistics* 6, pp. 3–73.

Dagum, Estela Bee and Silvia Bianconcini (2008). "The Henderson Smoother in Reproducing Kernel Hilbert Space". In: *Journal of Business & Economic Statistics* 26, pp. 536–545. URL: https://ideas.repec.org/a/bes/jnlbes/v26y2008p536-545.html.

Grun-Rehomme, Michel, Fabien Guggemos, and Dominique Ladiray (2018). "Asymmetric Moving Averages Minimizing Phase Shift". In: *Handbook on Seasonal Adjustment*. URL: ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.

Ladiray, Dominique (2018). "Moving Average Based Seasonal Adjustment". In: *Handbook on Seasonal Adjustment*. URL: ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.

Musgrave, John C (1964). "A set of end weights to end all end weights". In:

Pierce, David A. (Sept. 1980). "Data revisions with moving average seasonal adjustment procedures". In: *Journal of Econometrics* 14.1, pp. 95–114. URL: https://ideas.repec.org/a/eee/econom/v14y1980i1p95-114.html.

Proietti, Tommaso and Alessandra Luati (Dec. 2008). "Real time estimation in local polynomial regression, with application to trend-cycle analysis". In: *Ann. Appl. Stat.* 2.4, pp. 1523–1553. DOI: 10.1214/08-AOAS195. URL: https://doi.org/10.1214/08-AOAS195.

Wildi, Marc and Tucker McElroy (2019). "The trilemma between accuracy, timeliness and smoothness in real-time signal extraction". In: *International Journal of Forecasting* 35.3, pp. 1072–1084. URL: https://EconPapers.repec.org/RePEc:eee:intfor:v:35:y:2019:i:3:p:1072-1084.

Wildi, Marc and Tucker Mcelroy (Dec. 2013). "Optimal Real-Time Filters for Linear Prediction Problems". In: *Journal of Time Series Econometrics* 8. DOI: 10.1515/jtse-2014-0019.

Zellner, Arnold, Chansik Hong, and Chung-ki Min (1991). "Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques". In: *Journal of Econometrics* 49.1-2, pp. 275–304. URL: https://EconPapers.repec.org/RePEc:eee:econom:v:49:y:1991:i:1-2:p:275-304.