Alain QUARTIER-LA-TENTE

ENSAE 2ème année

Stage d'application

Année scolaire 2019 - 2020

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

2 Introduction

3 From seasonal adjustment to trend-cycle estimation

Indiquer pourquoi on se place dans le cadre de la désaisonnalisation.

Dire que l'on se place dans le cadre de X-12-ARIMA qui est une méthode non-paramétrique : on va s'intéresser à deux approches : Proietti and Luati (2008) (section 5) et Grun-Rehomme, Guggemos, and Ladiray (2018) (section 6).

Mentionner également Dagum (2018) (section @ref(sec:Dagum})) et Wildi (2018) / Wildi and McElroy (2019) (section 7).

Description du trilemma.

Parler aussi de la timeliness

4 Moving average and filters

A lot of papers describes the definition and the properties of moving average and linear filters (see for example Ladiray (2018)). Here 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 $\theta = (\theta_{-p},\dots,\theta_f)'$ such as:

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

- $p+f+1$ 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 $X_t = \mathrm{e}^{i\omega t}$, the result of the moving average M_θ in X_t is:

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

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

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

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[1]. For all symmetric moving average we have $\Phi_\theta(\omega) = 0$.

To sum up, applying a moving average to an harmonic times series affects 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[2].

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

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

The figure 1 illustrates the gain and the phase shift for $\omega = \pi/2$ and $X_t = \sin(\omega t)$.

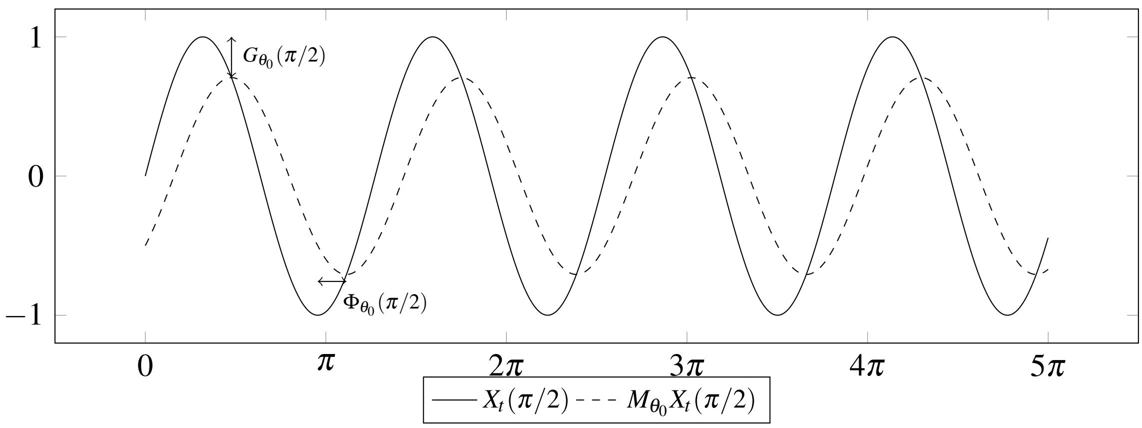

Figure 1: Smoothing of the time series $X_t = \sin(\omega t)$ by the moving average $M_{\theta_0} X_t = \frac{1}{2} X_{t-1} + \frac{1}{2} X_t$ for $\omega = \pi/2$.

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.

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

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

4.2.1 Trend preservation

Is 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.

- 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 conserve polynomial of degree $2d$ then it also conserve 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 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 $\mathbb{E}[\varepsilon_t] = 0\mathbb{E}$, $\mathbb{V}[\varepsilon_t] = \sigma^2$.

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

4.3 Real-time estimation and asymmetric moving average

For symmetric filters, the phase shift function is equal to zero. Therefore, there is no delay in any frequency: that's why they are prefered 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 summarised 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	$\sum_{k=-p}^{+f}\theta_k^2$
S_g	Smoothness (Guguemos)	$\sum_j(\nabla^3\theta_j)^2$
T_g	Timeliness (Guguemos)	$\int_0^{2\pi/12}\rho_\theta(\omega)\sin(\varphi_\theta(\omega))^2\,\mathrm{d}\omega$
A_w	Accuracy (Wildi)	$2\int_0^{2\pi/12}(\rho_s(\omega)-\rho_\theta(\omega))^2 h_{RW}(\omega)\,\mathrm{d}\omega$
T_w	Timeliness (Wildi)	$8\int_0^{2\pi/12}\rho_s(\omega)\rho_\theta(\omega)\sin^2\left(\frac{\varphi_\theta(\omega)}{2}\right)h_{RW}(\omega)\,\mathrm{d}\omega$
S_w	Smoothness (Wildi)	$2\int_{2\pi/12}^{\pi}(\rho_s(\omega)-\rho_\theta(\omega))^2 h_{RW}(\omega)\,\mathrm{d}\omega$
R_w	Residual (Wildi)	$8\int_{2\pi/12}^{\pi}\rho_s(\omega)\rho_\theta(\omega)\sin^2\left(\frac{\varphi_\theta(\omega)}{2}\right)h_{RW}(\omega)\,\mathrm{d}\omega$

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_g criteria from Wildi and McElroy (2019).
ρ_s represent the gain function of an Henderson symmetric filter to compare with the asymmetric ones.
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 arised 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. 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 observations $\left(y_{t+j}\right)_{j \in \llbracket -h,h \rrbracket}$. Then $\mu_t \simeq m_t$ with:

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

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

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

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

- the degree d of the polynomial;
- the number of neighbored $H = 2h+1$ (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:

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

where κ_j is a set of weights called *kernel*. We have $\kappa_j \geq 0$, $\kappa_{-j} = \kappa_j$ and with $K = diag(\kappa_{-h}, \dots, \kappa_h)$, 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[3]. Moreover, $X'w = e_1$ so:

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

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

5.1 Different kernels

In signal extraction, we generally look for weighting observations according to their distance from time t: this is the role of the kernel function. For that, we introduce a kernel function κ_j, $j = 0, \pm 1, \dots, \pm h$ with $\kappa_j \geq 0$ and $\kappa_j = \kappa_{-j}$. An important class of kernels is the Beta kernels. In the discrete, 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 encompass all kernels used in this report, except Henderson, trapezoidal and gaussian kernel.The following kernels are considered in this report:

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

- $r=1, s=0$ 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[4]

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

Henderson, trapezoidal and gaussian kernel are very specific:

- The kernel Henderson and trapezoidal function 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-12-ARIMA algorithm. Therefore it is never used to extract trend-cycle component.

The figure 2 summarises the coefficients of the different kernels. Analysing 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).

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

- 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.

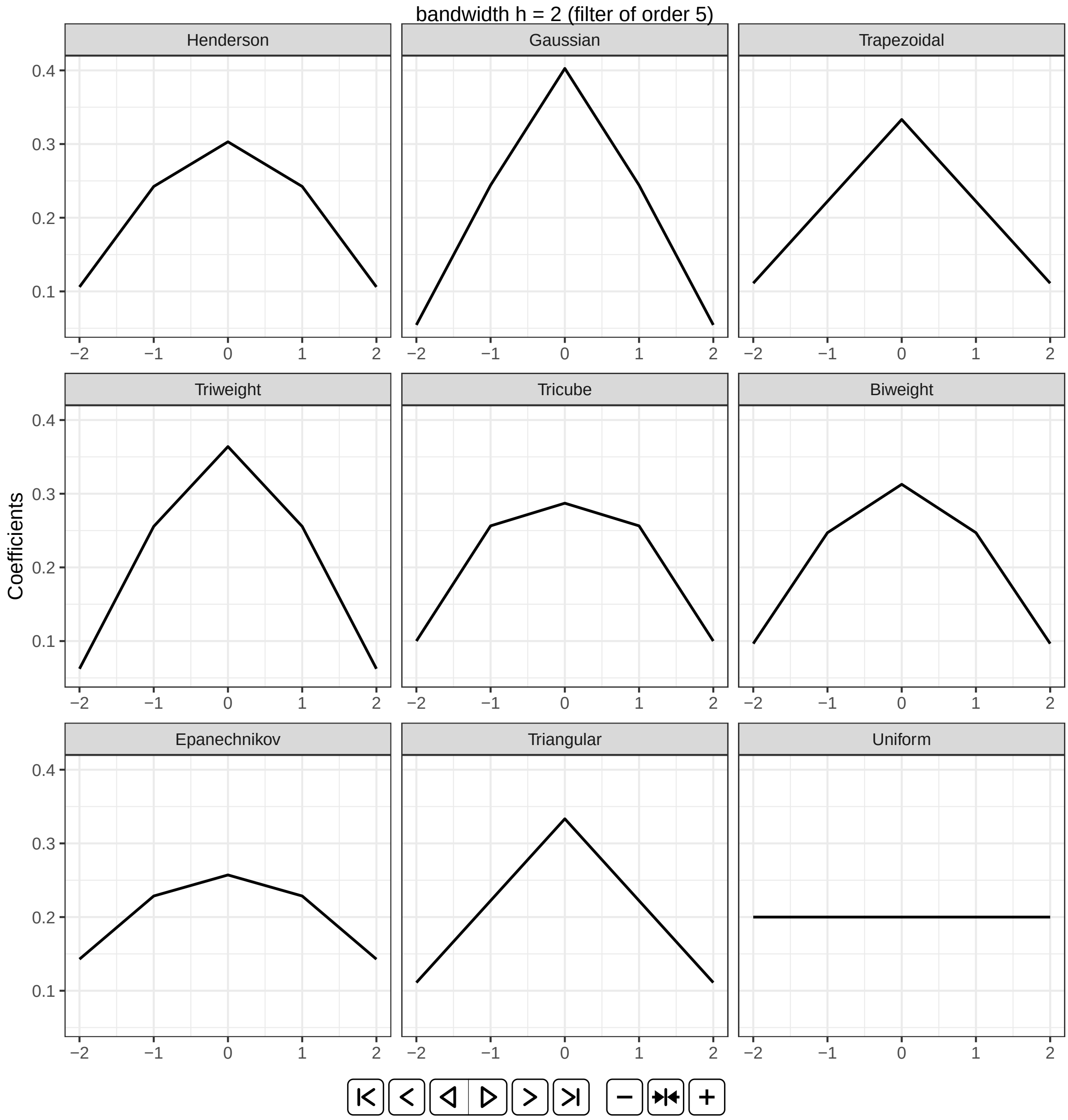

Figure 2: Coefficients of the different kernels for h from 2 to 30.

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 $h = 2$.

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 cubif fit ($p = 3$), widely used for trend estimation (for example it's the filter used in the seasonal adjustment software X-12-ARIMA). 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 *smothness 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 differents kernels presented in different kernels presented in section 5.1 and for different bandwidth 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 interprated 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.

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

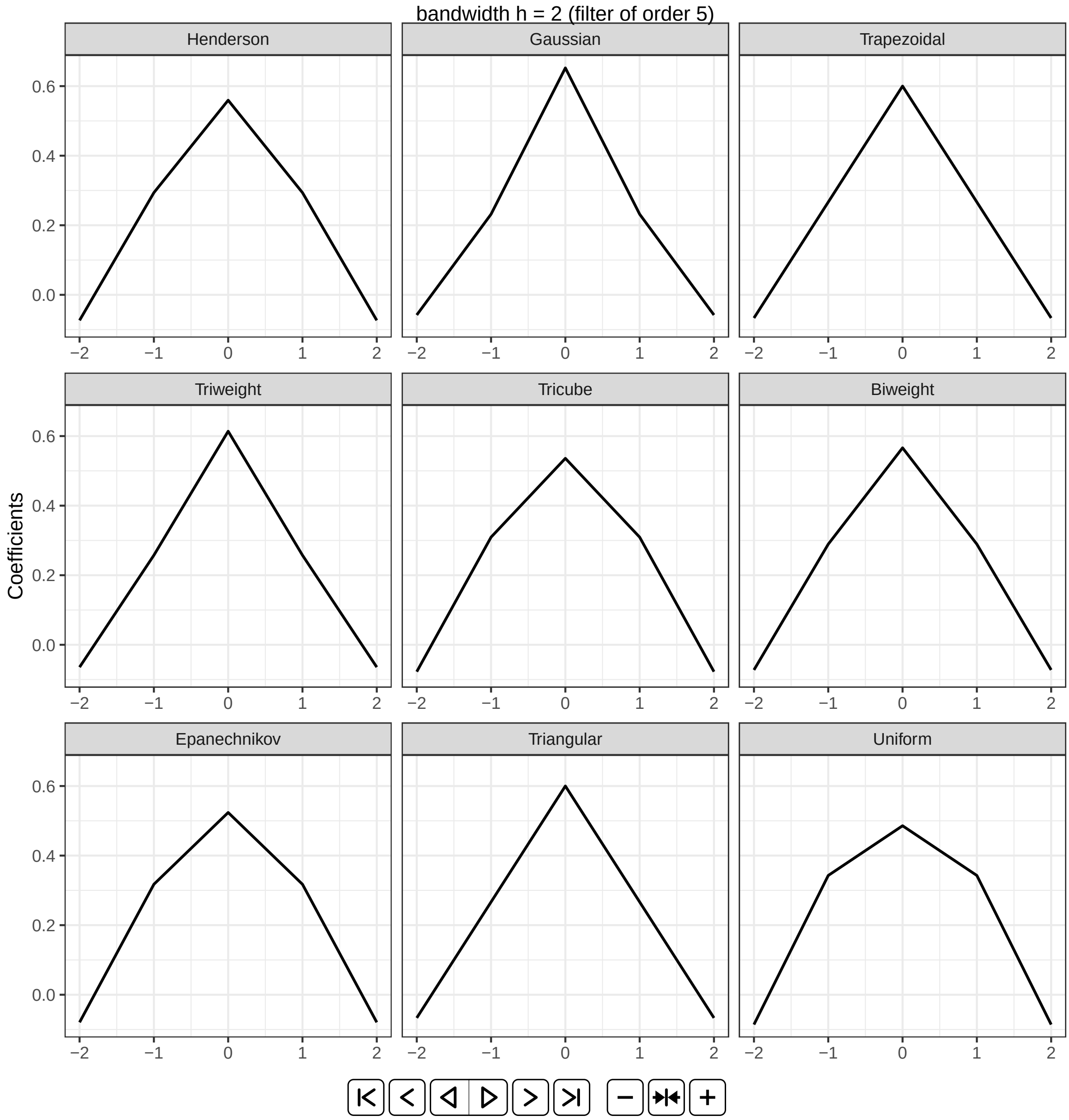

Figure 3: Coefficients of symmetric filters computed by local polynomial of degree 3, according to the differents kernels and for h from 2 to 30.
Note: to see the animation the PDF must be open with Acrobat Reader, KDE Okular, PDF-XChange or Foxit Reader. Otherwise you will only be able to see the results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

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).

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

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 correspond 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.

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

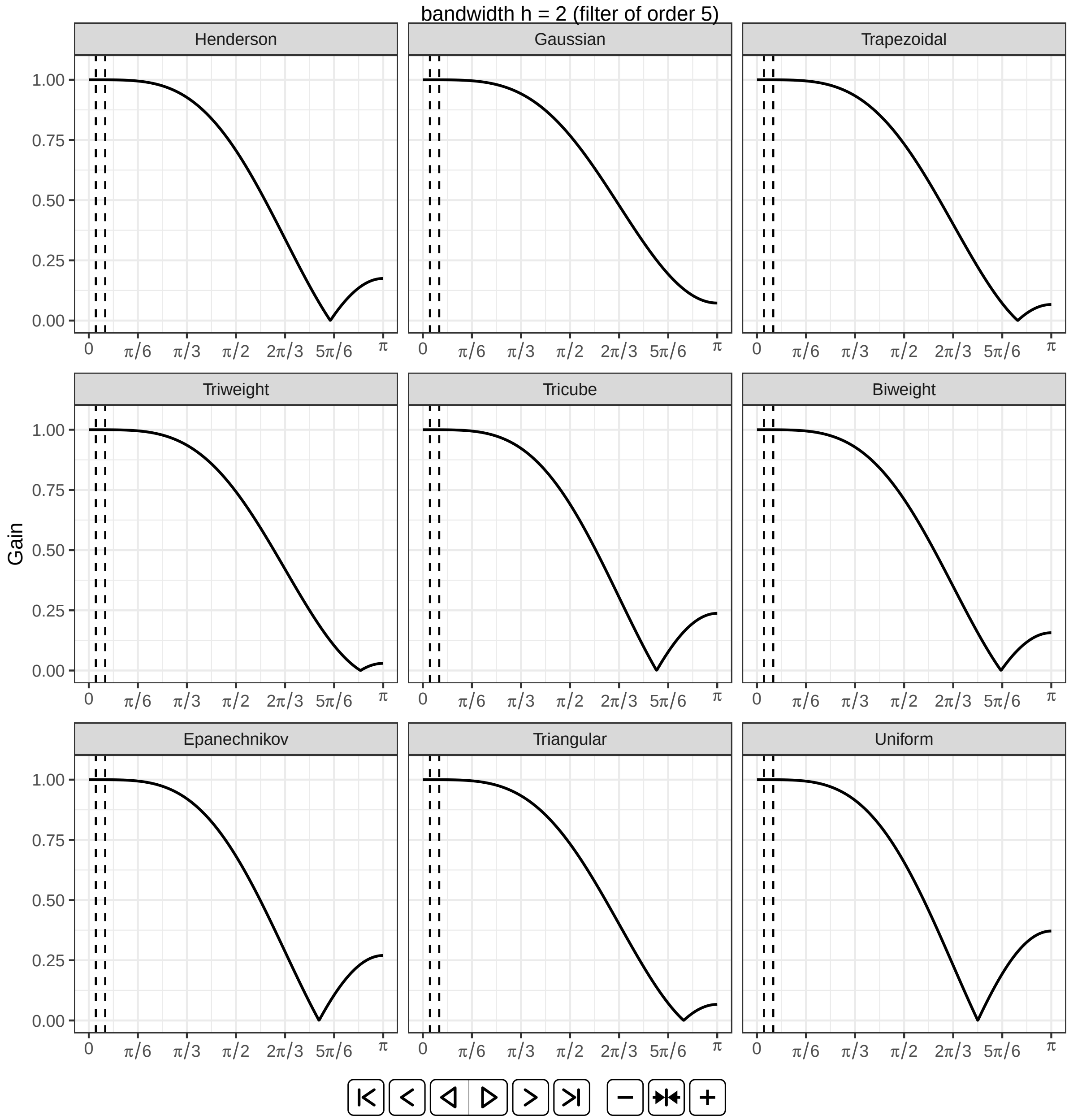

Figure 4: Gain functions from 0 to $2\pi/12$ of symmetric filters computed by local polynomial of degree 3, according to the differents kernels and for h from 2 to 30.
Note: the two horizontal lines corresponds to the frequencies $2\pi/84$ (cycle of 7 years) and $2\pi/36$ (cycle of 3 years). To see the 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 $h = 2$.

1. Build a 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 a 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).

Rewritting 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 satisfy 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 have 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.

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.

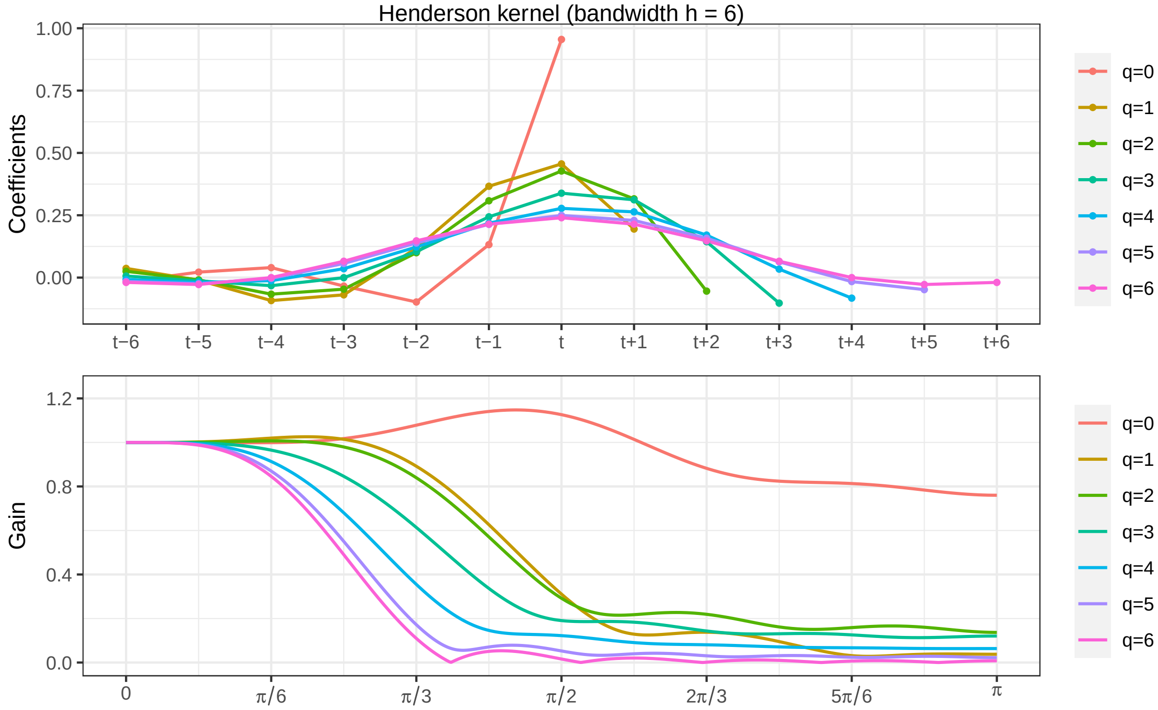

Figure 5: Coefficients and gain function of direct asymmetric filters (DAF) computed by local polynomial of degree 3 with the Henderson kerne for $h=6$.

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:

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

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 (see ajouter biblio). With the 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[5] is used to determine the bandwidth to used for the Henderson filter. For monthly data:

- if $R < 1$ a 9-term Henderson is used ($h = 4$);
- if $1 \leq R \leq 3.5$ a 13-term Henderson is used ($h = 6$);
- if $3.5 < R$ a 23-term Henderson is used ($h = 12$).

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 ($d^* = 0$), v depend on the ratio δ_1^2/σ^2. We obtain Musgrave filters when the Henderson kernel is used.
2. *Quadratic-Linear* (QL): y_t quadratic ($d = 2$) and v preserve linear signals ($d^* = 1$), v depend on the ratio δ_2^2/σ^2.
3. *Cubic-Quadratic* (CQ): y_t cubic ($d = 3$) and v preserve linear signals ($d^* = 1$), v depend on the ratio δ_2^2/σ^2.

parler du tableau 3

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

Summary — local polynomial filters

Advantages:

Drawbacks:

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

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	2.705	1.796
QL	0	0.000	-0.473	0.711	5.149	0.047	0.067	1.894	1.054	0.967
CQ	0	0.000	0.000	0.913	11.942	0.015	0.016	2.231	0.997	0.573
DAF	0	0.000	0.000	0.943	14.203	0.003	0.015	2.178	1.005	0.571
$q=1$										
LC	0	-0.121	-0.525	0.268	0.433	4.797	0.009	0.119	1.527	0.832
QL	0	0.000	-0.061	0.287	0.707	0.694	0.005	0.192	1.134	0.613
CQ	0	0.000	0.000	0.372	0.571	0.158	0.022	0.575	1.064	0.785
DAF	0	0.000	0.000	0.409	0.366	0.061	0.020	0.760	1.044	0.717
$q=2$										
LC	0	0.003	1.076	0.201	0.080	0.347	0.009	0.012	1.047	0.424
QL	0	0.000	0.033	0.215	0.052	2.083	0.000	0.011	1.230	0.638
CQ	0	0.000	0.000	0.370	0.658	0.131	0.021	0.558	1.059	0.755
DAF	0	0.000	0.000	0.398	0.768	0.023	0.017	0.677	1.030	0.634

6 General optimisation problem: Grun-Rehomme, Guggemos, and Ladiray (2018)

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" become we want the output signal to be as close as possible to the input signal where the noise component is removed
$$F_g(\theta) = \sum_{k=-p}^{+f} \theta_k^2$$
- *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.
$$S_g(\theta) = \sum_j (\nabla^3\theta_j)^2$$
- *Timeliness*, T_g: it mesures 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.

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

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 $f\colon (\rho,\varphi) \mapsto \rho^2 \sin(\varphi)^2$. Indeed, for this function, the timeliness criterion is analytically solvable ($T_g = \theta' T \theta$ with T a square 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.

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

The Henderson symmetric filters can for example be computed with

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

Summary — FST filters

Advantages:

Drawbacks:

7 Data-dependent filter: Wildi and McElroy (2019)

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.

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:

$$\mathbb{E}\left[(y_t-\hat{y}_t)^2\right]=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\Gamma_s(\omega)-\Gamma_\theta(\omega)\right|^2h(\omega)\,\mathrm{d}\omega=\frac{1}{2\pi}\times2\times\int_{0}^{\pi}\left|\Gamma_s(\omega)-\Gamma_\theta(\omega)\right|^2h(\omega)\,\mathrm{d}\omega \tag{1}$$

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)).

$$\begin{aligned}\left|\Gamma_s(\omega)-\Gamma_\theta(\omega)\right|^2&=\rho_s(\omega)^2+\rho_\theta(\omega)^2+2\rho_s(\lambda)\rho_\theta(\lambda)\left(1-\cos(\varphi_s(\omega)-\varphi_\theta(\omega)\right)\\&=\left(\rho_s(\omega)-\rho_\theta(\omega)\right)^2+4\rho_s(\lambda)\rho_\theta(\lambda)\sin^2\left(\frac{\varphi_s(\omega)-\varphi_\theta(\omega)}{2}\right)\end{aligned} \tag{2}$$

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 (1) can then be decomposed additively into four quantities:

$$\begin{aligned}Accuracy=A_w&=2\int_0^{\omega_1}\left(\rho_s(\omega)^2-\rho_\theta(\omega)\right)^2h(\omega)\,\mathrm{d}\omega\\Smoothness=S_w&=8\int_0^{\omega_1}\rho_s(\omega)\rho_\theta(\omega)\sin^2\left(\frac{\varphi_s(\omega)-\varphi_\theta(\omega)}{2}\right)h(\omega)\,\mathrm{d}\omega\\Timeliness=T_w&=2\int_{\omega_1}^{\pi}\left(\rho_s(\omega)^2-\rho_\theta(\omega)\right)^2h(\omega)\,\mathrm{d}\omega\\Residual=R_w&=8\int_{\omega_1}^{\pi}\rho_s(\omega)\rho_\theta(\omega)\sin^2\left(\frac{\varphi_s(\omega)-\varphi_\theta(\omega)}{2}\right)h(\omega)\,\mathrm{d}\omega\end{aligned}$$

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 the authors suggest to compute then minimizing a weighted sum of the first three indicators:

$$\mathscr{M}(\vartheta_1,\vartheta_2)=\vartheta_1T_w(\theta)+\vartheta_2S_w(\theta)+(1-\vartheta_1-\vartheta_2)A_w(\theta)$$

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 usefull 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, we take for the symmetric filter the Henderson filter and we fix the spectral density to the one of a random walk:

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

The formula presented in table 1 are then found using the fact that for a symmetric filter the phase shit is equal to zero ($\varphi_s(\omega) = 0$).

Summary — Data-dependent filters

Advantages:

Drawbacks:

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 $\mathbb{L}^2(f_0)$[6], 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 weight of a symmmetric filter are then derived with:

$$\forall j \in \llbracket -m,m \rrbracket : w_j = \frac{K_p(j/b)}{\sum_{i=-m}^{m} K_p(i/b)}$$

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

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

where b is a time-invariant global bandwidth parameter.

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

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

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

In Bee Dagum and Bianconcini (2015), the authors suggest to perform an optimal bandwidth selection (parameter b), decomposing the mean squared revision error as in equation (2). Therefore, the bandwidth can for example be chosen to minimise the mean squared revision error, the phase shift, etc.

$$b_{q,G} = \min_{b_q} \sqrt{2\int_0^{1/2} \left(\rho_s(\omega) - \rho_\theta(\omega)\right)^2 \mathrm{d}\omega}$$

$$b_{q,\gamma} = \min_{b_q} \sqrt{2\int_0^{1/2} |\Gamma_s(\omega) - \Gamma_\theta(\omega)|^2 \mathrm{d}\omega}$$

$$b_{q,\varphi} = \min_{b_q} \sqrt{2\int_0^{1/2} \rho_s(\lambda)\rho_\theta(\lambda)\sin^2\left(\frac{\varphi_\theta(\omega)}{2}\right)}$$

Summary — RKHS filters

Advantages:

Drawbacks:

9 Comparison of the different filters

Mettre les graphiques 3D,

https://aqlt.shinyapps.io/FSTfilters/

10 Results

6:

- almost same symmetric trend for all the filters: small differences during "turning points". Differences less than one point for between all the kernels, uniform and trapezoidal more "atypical", Differences less than 0,5 point for the other kernels. Différences p/r Henderson toujours dans le même sens sauf pour triweight (?) : sous Henderson pendant les "cycles bas", au-dessus pendant "cycles hauts"
- revisions error are higher with uniform and trapezoidal kernels, lower with triweight, regardless of the asymmetric filter.
- in mean, the variance revision error is always lower with the LC filter and higher with the DAF. However, during turning-points (for example around the 2008 crisis), the LC filter have a higher bias than the other filters. For the French IPI, the CQ method seems to give the better results around the 2008 crisis (analyzing mean squared error revision between 2007 and 2011). However, reproducing the same procedure to the IPI of other countries of the European Union (and the United-Kingdom), we find that for most of the series (23 of 28) its the QL filter that gives the better results during the 2008 crisis.

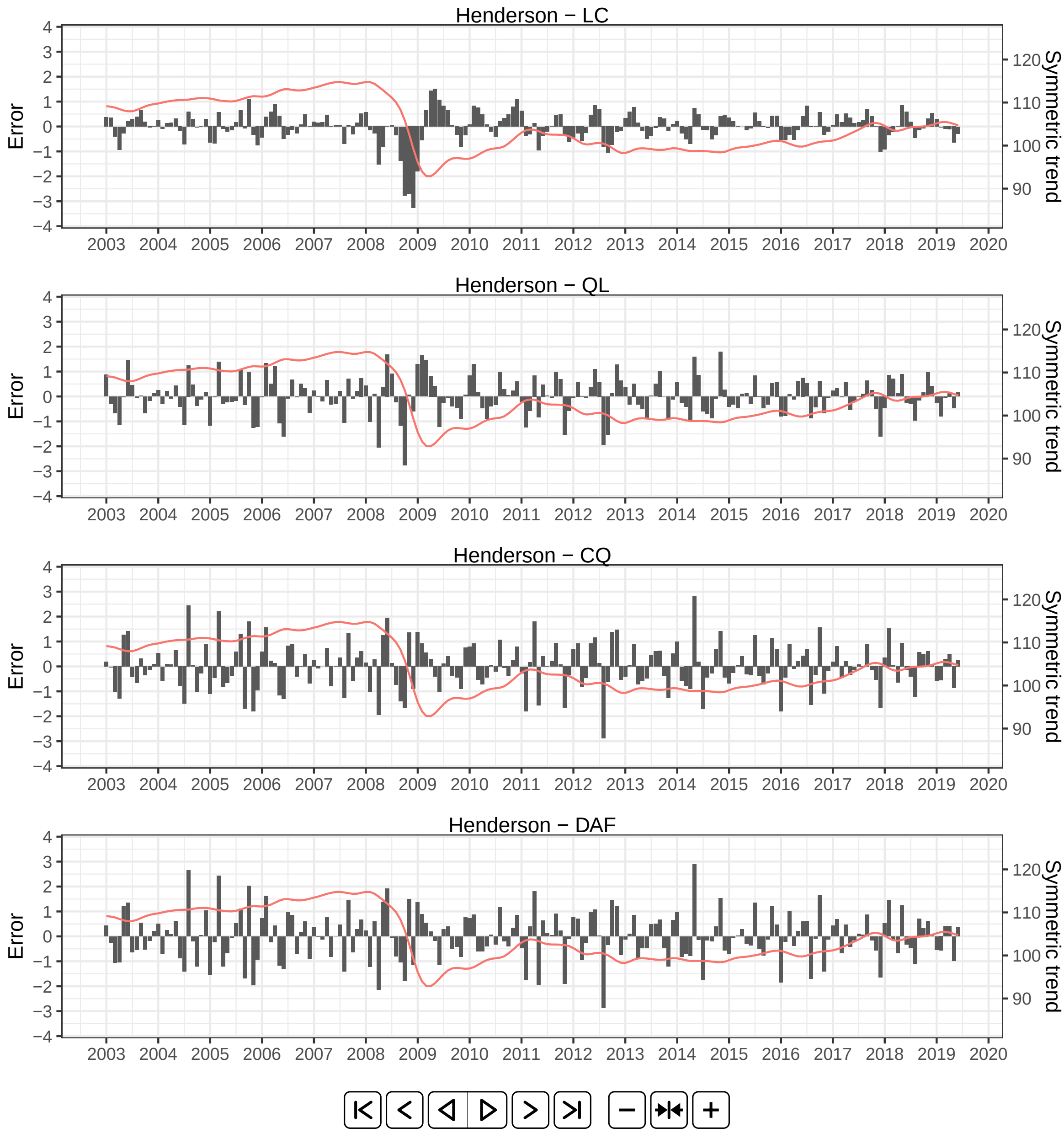

Figure 6: Revision error of the different real-time ($q = 0$) asymmetric filters for $h = 6$, for the IPI-FR.

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 the Henderson kernel.

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.

Dagum, Estela Bee (2018). "Trend-Cycle Estimation". In: *Handbook on Seasonal Adjustment*. URL: ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.

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.

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 (2018). “Real Time Trend Extraction and Seasonal Adjustment”. In: *Handbook on Seasonal Adjustment*. URL: ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.

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.