Modèle de page de couverture du rapport à respecter

NOM Prénom de l'élève

ENSAE 2ème année

Stage d'application ou stage long[1]

Année scolaire XXXX-XXXX

Titre du rapport

Mention « confidentiel » si besoin

Nom de l'entreprise

Ville (ou Ville et Pays)

Maître de stage : Prénom NOM

Date de début à date de fin de stage

[1] *Indiquer la mention correspondant à votre cas.*

ENSAE Paris

TSA 26644

www.ensae.fr

Service des relations entreprises et des stages

5, avenue Henry Le Chatelier- 91764 PALAISEAU CEDEX - FRANCE - Tél : +33 (0)1 70 26 67 39 - Courriel : stage@ensae.fr

Contents

1 Introduction

2 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)). He 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:

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

- $p+f+1$ 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 *bandwith*.

2.1 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 $\Gamma_\theta(\omega)=\sum_{k=-p}^{+f}\theta_k e^{i\omega k}$ is called the *transfer 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].

[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.

2.2 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.

2.2.1 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:

$$\begin{cases} \sum_{k=-p}^{+f} \theta_k & = 1 \\ \sum_{k=-p}^{+f} k\theta_k & = 0 \end{cases}$$

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

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

2.2.2 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$\mathbb{E}\left[\varepsilon_t\right] = 0\mathbb{E}$, $\mathbb{V}\left[\varepsilon_t\right] = \sigma^2$.

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

2.3 Asymmetric moving average

2.4 Quality indicators (? trouver un autre nom)

To compare the different moving average / filters the following indicators are used:

$$\begin{cases} \text{constant bias} = \text{bias0} & = \sum_{k=-p}^{+f} \theta_k \\ \text{linear bias} = \text{bias1} & = \sum_{k=-p}^{+f} k\theta_k \\ \text{quadratic bias} = \text{bias2} & = \sum_{k=-p}^{+f} k^2\theta_k \\ \text{variance reduction} & = \sum_{k=-p}^{+f} \theta_k^2 \end{cases}$$

à compléter

3 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 $\left(y_{t+j}\right)_{j \in [\![-h,h]\!]}$. 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:

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

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 *bandwith* 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:

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

where κ_j is a set of weights called *kernel*. We have $\kappa_j \geq 0$, $\kappa_{-j} = \kappa_j$ and with $K = diag(\kappa_{-h}, \dots, \kappa_h)$, 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.

[3]Due to the symmetry of the kernel weights κ_j.

3.1 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 $r>0$, $s\geq 0$. 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:
$$\kappa_j^{TC} = \left(1-\left|\frac{j}{h+1}\right|^3\right)^3$$
- Gaussian kernel[4]:
$$\kappa_j^G = \exp\left(-\frac{j^2}{4h}\right)$$
- Henderson kernel (see section 3.1.1 for its construction):
$$\kappa_j = \left[1-\frac{j^2}{(h+1)^2}\right]\left[1-\frac{j^2}{(h+2)^2}\right]\left[1-\frac{j^2}{(h+3)^2}\right]$$
- Trapezoidal kernel:
$$\kappa_j^{TP} = \begin{cases} \frac{1}{3(2h-1)} & \text{if } j=\pm h \\ \frac{2}{3(2h-1)} & \text{if } j=\pm(h-1) \\ \frac{1}{2h-1} & \text{otherwise} \end{cases}$$

The figure 1 summarises the coefficients of the different kernels. Analysing the coefficients we can already anticipate some properties of the associated filters:

[4] The gaussian kernel is generally defined as $\mathrm{e}^{-\frac{x^2}{2\sigma^2}}$. Here we take $\sigma^2 = 2/h$ with the following idea: the biggest h is, the narrowest the filter is.

- For h small the triweight kernel has the narrowest distribution, for h high ($h \geq 15$) the gaussian kernel become narrowest. 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.

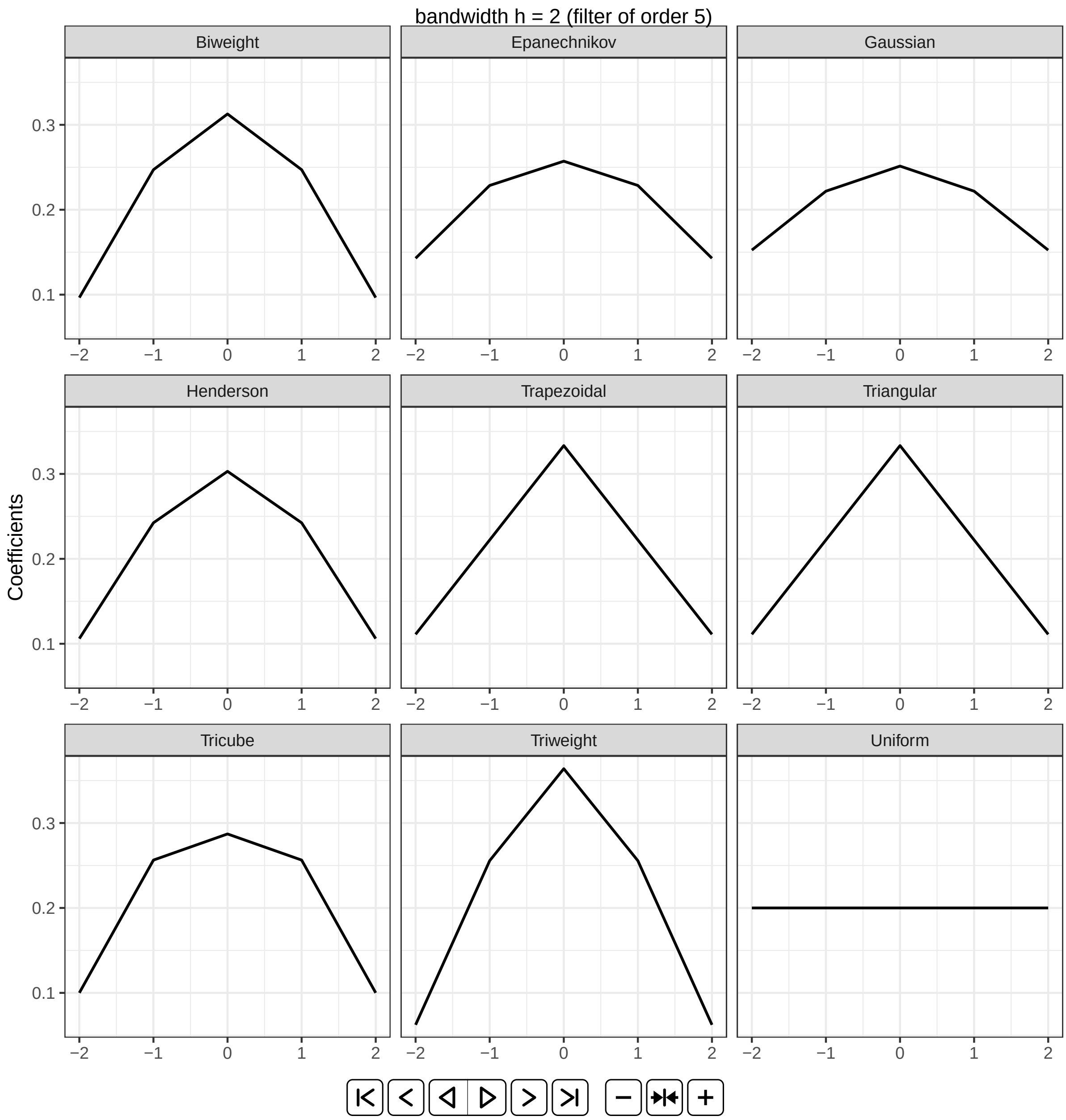

Figure 1: Coefficients of the different kernels for h from 2 to 30.

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 $h = 2$.

3.1.1 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 ($p = 3$). It is often used Henderson for trend estimation (for example it's the filter used in the seasonal adjustment). For a fixed bandwith, 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 criteria*: $S = \sum_j (\nabla^3 \theta_j)^2$;

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 3.1.

3.1.2 Analysis of symmetric filters

In this section, all the filters are computed by local polynomial of degree $d = 3$. The figure 2 plots the coefficients of the filters for the differents kernels presented in different kernels presented in section 3.1 and for different bandwith h. The table 1 shows the variance reduction of the different filters. We find the similar results than in section 1:

- For h small the triweight kernel gives the filter with the narrowest distribution, for h high ($h \geq 15$) the gaussian kernel become narrowest. 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 high the Henderson filter is equivalent to the one computed by the triweight kernel.

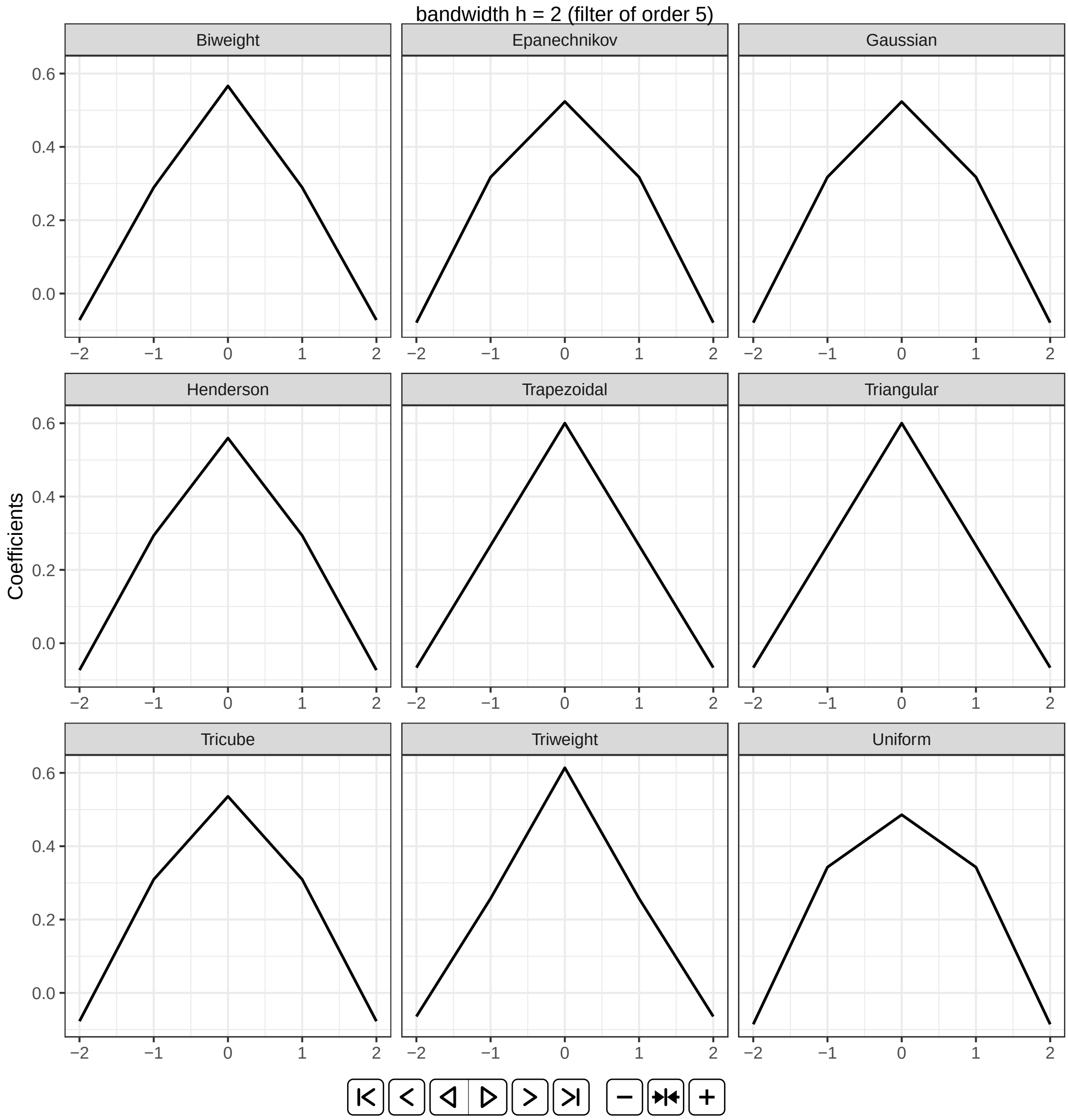

Figure 2: Coefficients of symmetric filters computed by local polynomial of degree 3, according to the differents kernels and for h from 2 to 30.

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 $h = 2$.

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 higher coefficients (they are more 80% smaller than the central coefficient for all kernels, except for uniform and trapezoidal with high bandwith).

Table 1: 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

3.1.3 Gain functions

Figure 3 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). That's why we are only interested in low frequencies. 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 $2\pi/84$.

When the bandwith 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).

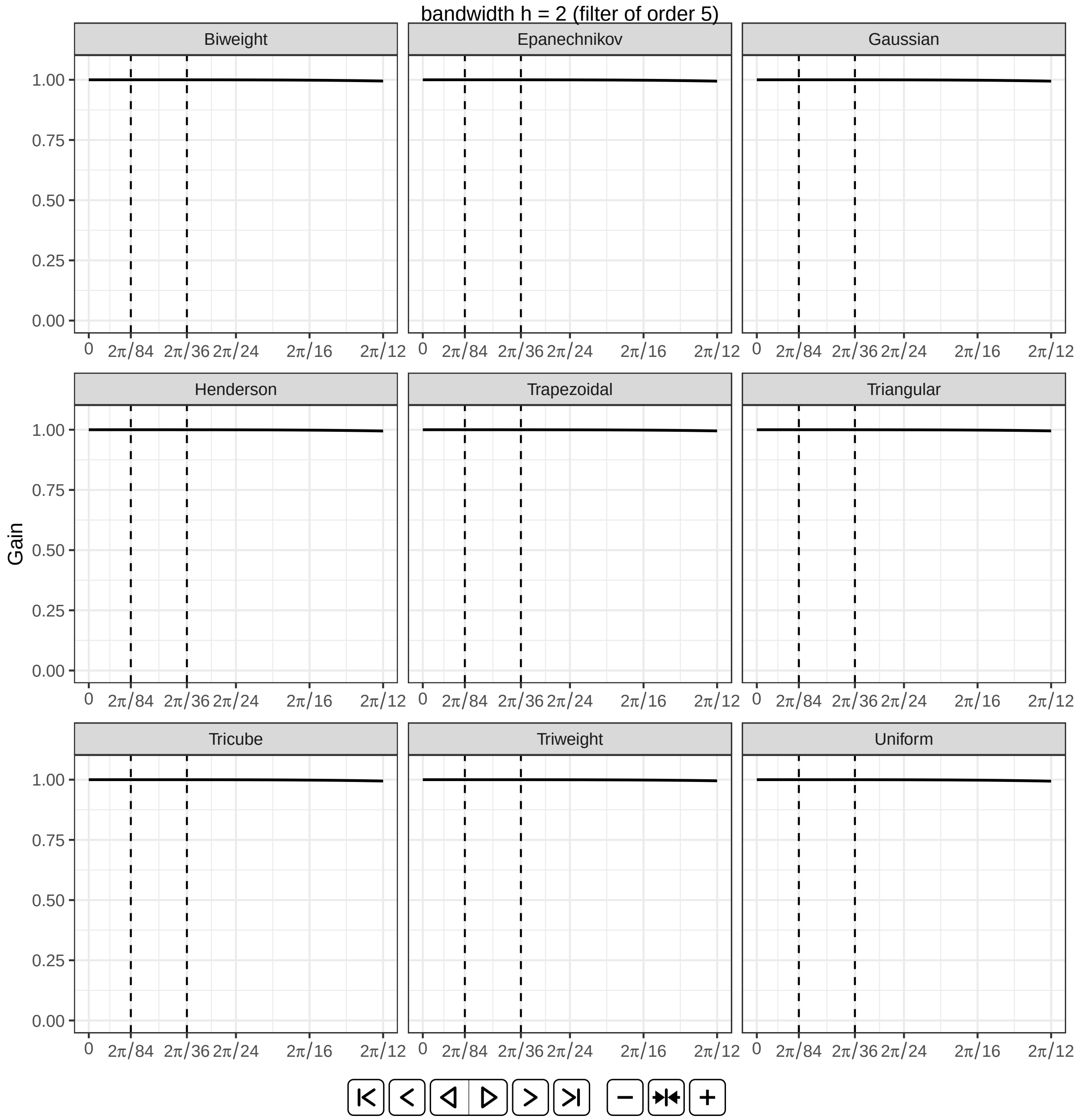

Figure 3: Gain functions from 0 to $2\pi/12$ of symmetric filters computed by local polynomial of degree 3, according to the differents kernels and for h from 2 to 30.

Note: 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 $h = 2$.

References

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

Proietti, Tommaso, and Alessandra Luati. 2008. "Real Time Estimation in Local Polynomial Regression, with Application to Trend-Cycle Analysis." *Ann. Appl. Stat.* 2 (4): 1523–53. https://doi.org/10.1214/08-AOAS195.