ESTUDIOS DE ECONOMÍA APLICADA VOL. 28-3 2010 PÁGS. 553-574

Real Time Analysis Based on Reproducing Kernel Henderson Filters

SILVIA BIANCONCINI
Department of Statistics
UNIVERSITY OF BOLOGNA, ITALIA

BENOIT QUENNEVILLE
Time Series Research and Analysis Center
BUSINESS SURVEY METHODS DIVISION, STATISTICS CANADA

e-mail: silvia.bianconcini@unibo.it; Benoit.Quenneville@statcan.gc.ca

ABSTRACT

Recently, reproducing kernel Hilbert spaces have been introduced to provide a common approach for studying several nonparametric estimators used for smoothing functional time series data (Dagum and Bianconcini, 2006 and 2008). The reproducing kernel representation is based on the derivation of the density function (i.e. a second order kernel) embedded on the linear filter. This is the starting point for deriving higher order kernels, which are obtained from the product of the density and its orthonormal polynomials. This paper focuses on the Henderson filter, for which two density functions and corresponding hierarchies have been derived. The properties of the Henderson reproducing kernels are analyzed when the filters are adapted at the end of the sample period. The optimality criterion satisfied as well as the influence of the kernel order and bandwidth parameter are studied.

Keywords: Henderson Filters, Kernel Hilbert Spaces, Smoothing Estimators.

Análisis en tiempo real basado en la reproducción de los filtros de núcleo de Henderson

RESUMEN

Recientemente, la reproducción de los espacios de núcleo (kernel) de Hilbert se ha ido extendiendo con el objetivo de proporcionar un enfoque común para estudiar diversos estimadores no paramétricos de alisado para datos de series temporales funcionales (Dagum y Bianconcini, 2006 y 2008). La representación del núcleo reproducido se basa en la obtención de la función de densidad (i.e., un núcleo (kernel) de segundo orden) incluido en el filtro linear. Esto constituye el punto de partida para derivar núcleos de orden superior, obtenidos a partir del producto de la densidad por sus polinomios ortonormales. Este artículo se centra en el filtro de Henderson, para el que se obtienen dos funciones de densidad y sus correspondientes jerarquías. Se analizan las propiedades de los núcleos de Henderson reproducidos cuando se adaptan los filtros al final del período muestral. Además, se estudia el criterio de optimalidad que satisfacen así como la influencia del orden del núcleo y del parámetro del ancho de banda.

Palabras clave: Filtro de Henderson, espacios de núcleo de Hilbert, estimadores de alisado.

Clasificación JEL: C30, C32.

Artículo recibido en agosto de 2010 y aceptado en septiembre de 2010.
Artículo disponible en versión electrónica en la página www.revista-eea.net, ref. ə-28313.

1. INTRODUCTION

The linear smoother developed by Henderson (1916) is the most widely applied to estimate the trend-cycle latent component in nonparametric seasonal adjustment software, such as the U.S.Bureau of the Census X11 method (Shiskin et al. 1967) and its variants, the X11ARIMA (Dagum 1988) and X12ARIMA (Findley *et al.* 1998). The study of the properties and limitations of the Henderson filter has been done in different contexts and attracted the attention of a large number of authors. Among others, Dagum (1996), Gray and Thomson (1996), Doherty (2001), Quenneville, Ladiray and Lefrançois (2003), and Dagum and Bianconcini (2006 and 2008) focused on the properties of the corresponding asymmetric filters applied to the most recent observations. The latter are closely analyzed by users in order to assess the stage at which the economy stands or, more generally, to evaluate the most recent developments in the phenomenon under investigation. Hence, a problem that is common to all the methodologies is the reliability of these trend estimates. At the end of the sample period, the data are subject to revisions when additional observations are made because of the use of asymmetric filters, hence these filters have to be optimal in some way. The concern of the paper is real time estimation of the underlying trend in a time series by means of asymmetric weights associated to the symmetric Henderson filter. Real time estimation is of outmost importance in many fields, such as economics, and deals with the estimation of a signal at time t using the observations available up to and including time t.

As a local polynomial estimator, the Henderson asymmetric filters can be derived through automatic adaptations at the end of the sample period. This implies that the bias and the variance near the boundary are of the same order of magnitude as in the interior (see e.g. Fan and Gijbels, 1996). It turns out that the variance inflation resulting from the one-sided real time Henderson filter is very high, and that the filter is strongly localized at the current observations. In order to overcome such limitations, alternative strategies have been applied in the literature.

The idea embodied in the X11ARIMA seasonal adjustment procedure is to apply the symmetric two-sided Henderson filter to the series extended by forecasts (Dagum, 1982). This strategy is safer, provided that we are capable of producing optimal forecasts according to some time series model of the ARIMA class. An intuitive and easily fact is that if the forecasts are optimal in the mean square error sense, then the variance of the revision is a minimum (Wallis, 1983).

In X11 based seasonal adjustment procedures, the end-point estimates of the trend-cycle are obtained using asymmetric filters that were developed by Musgrave (1964), specifically to minimize revisions for a certain class of time series. Doherty (2001) is the most complete study on how Musgrave surrogates, *i.e.* the X11 asymmetric trend-cycle averages, are computed. Quenneville and Ladiray (2000) also studied the dependence of these surrogates on a parameter, that is restricted to predetermined values in X11. They found that the differences in terms of revisions between the Musgrave averages with the estimated parameter and the Henderson average with forecasts from a simple ARIMA model can be very small.

Even if Musgrave asymmetric filters are optimal in terms of minimum mean square revision error of the estimates, they are derived following a different optimization criteria with respect to the symmetric Henderson filter, with the consequence that the asymmetric filters do not converge monotonically to the symmetric one. At this regard, Dagum and Bianconcini (2006 and 2008) provided a different characterization of the Henderson weights within the Reproducing Kernel Hilbert Space (RKHS) methodology. According to this approach, a continuous kernel representation of the Henderson filter is obtained, and the same kernel function is used to derive both symmetric and asymmetric weights. The density function (*i.e.* a second order kernel) embedded on the linear filter is firstly determined. This is the starting point for obtaining higher order kernels, which are based on the product of the density and its orthonormal polynomials. For each kernel order, the asymmetric filters can be derived coherently with the corresponding symmetric weights. In the particular case of the currently applied asymmetric Henderson filters, those obtained by means of the RKHS are shown to have superior properties from the viewpoint of signal passing, noise suppression, and revisions.

In this paper, we analyze the statistical properties of these asymmetric filters, paying particular attention on the selection of the kernel function and bandwidth parameter. Section 2 discusses the reproducing kernel representations of the classical Henderson symmetric filter as based on three different density functions. We illustrate the properties of the third order kernels within each hierarchy, and evaluate the problem of the selection of optimal bandwidths in the kernel functions used to derive the symmetric Henderson weights. Section 3 analyzes the statistical properties of the boundary Henderson kernels, obtained by applying the so called "cut-and-normalize" method, and of the corresponding asymmetric weights. Finally, Section 5 gives the conclusions.

2. KERNEL REPRESENTATIONS OF THE HENDERSON FILTER

Let $\{y_t, t = 1, 2, ..., N\}$ denote the input series, that is assumed to be decomposed into the sum of a systematic component, called the signal (or nonstationary mean) g_t , plus an erratic component u_t, called the noise, such that.

$$y_t = g_t + u_t \tag{1}$$

The noise u_t is assumed to be either a white noise, $WN(0, \sigma_u^2)$, or, more generally, to follow a stationary and invertible AutoRegressive Moving Average (ARMA) process.

If $\{y_t, t = 1, 2, ..., N\}$ is seasonally adjusted or without seasonality, the signal g_t represents the trend and cyclical components, usually referred to as trend-cycle for they are estimated jointly.

The trend-cycle can be deterministic or stochastic, and have a global or a local representation. It can be represented *locally* by a polynomial of degree p of a variable j, which measures the distance between y_t and the neighboring observations y_{t+j} . Equivalently, the trend-cycle estimate $\hat{g}_t$ consists in a weighted average applied in a moving manner (Kendall, Stuart and Ord 1983), such that

$$\hat{g}_t = \sum_{i=-m}^{m} w_j y_{t+j} \tag{2}$$

where w_j, $j < N$, denotes the weights to be applied to the observations y_{t+j} to get the estimate $\hat{g}_t$ for each point in time $t = 1, 2, ..., N$. Several nonparametric estimators, based on different sets of weights w_t, $j = -m, ..., m$, have been developed in the literature.

Recognition of the fact that the smoothness of the estimated trend-cycle curve depends directly on the smoothness of the weight diagram led Henderson (1916) to develop a formula which makes the sum of squares of the third differences of the smoothed series a minimum for any number of terms. Henderson's starting point was the requirement that the filter should reproduce a cubic polynomial trend without distortion. He proved that three alternative smoothing criteria give the same formula, as shown explicitly by Kenny and Durbin (1982) and Gray and Thomson (1996): (1) minimization of the variance of the third differences of the series defined by the application of the moving average (min var (Δ^3 $\hat{y}_t$)); (2) minimization of the sum of squares of the third differences of the coefficients of the moving average formula (min $\sum_{j=-m}^{m}=(\Delta^3 w_j)^2)$; (3) fitting a cubic polynomial by weighted least squares to the observations $y_{t+j}, j=-m,...,m$, where the weights are chosen as to minimize the sum of squares of their third differences. Representing the latter by W_j,$j = -m, ..., m$, where $W_j = W_{-j}$, the problem is the minimization of

$$\sum_{j=m}^{m} W_j [y_{t+j} - a_0 - a_1 j - a_2 j^2 - a_3 j^3]^2 \tag{3}$$

where the solution for the constant term $\hat{a}_0$ is the smoothed observation $\hat{g}_t$.. Henderson (1916) showed that $\hat{g}_t$ is given by

$$\hat{g}_t = \sum_{j=-m}^{m} \varphi(j)\, W_j y_{t+j} \tag{4}$$

where $\varphi(j)$ is a cubic polynomial whose coefficients have the property that the smoother reproduces the data if they follow a cubic. Henderson also proved the converse: if the coefficients of a cubic-reproducing summation formula $\{w_j, j = -m, ..., m\}$ do not change their sign more than three times within the filter span, then the formula can be represented as a local cubic smoother with weights W_j >0 and a cubic polynomial $\varphi(j)$, such that $\varphi(j)\ W_j = w_j$.

Henderson (1916) measured the amount of smoothing of the input series by $\sum (\Delta^3\ y_t)^2$ or equivalently by the sum of squares of the third differences of the weight diagram, $\sum (\Delta^3\ w_j)^2$. The solution is that resulting from the minimization of a cubic polynomial function by weighted least squares with

$$W_j \propto ((m+1)^2 - j^2)\,((m+2)^2 - j^2)\,((m+3)^2 - j^2) \tag{5}$$

as the weighting penalty function of criterion (3) above. Following Henderson (1916), the weight diagram $\{w_j, j = -m, ..., m\}$ corresponding to eq. (5), known as Henderson's ideal formula, is obtained, for a filter length equal to $2m' - 3$, by

$$w_j = \frac{315\,((m'+1)^2 - j^2)\,(m'^2 - j^2)\,((m'+1)^2 - j^2))\,(3m'^2 - 16 - 11j^2)}{8m'(m'^2-1)\,(4m'^2-1)\,(4m'^2-9)\,(4m'^2-25)} \tag{6}$$

This optimality result has been rediscovered several times in modern literature, usually for asymptotic kernel variants. At this regard, Loader (1999) showed that the Henderson's ideal formula (6) is a finite sample variant of a kernel with second order vanishing moments which minimizes the third derivative of the function given by Müller (1984). In particular, Loader showed that for large m, the weights of Henderson's ideal penalty function W_j are approximately $m^6\ W(j/m)$, where $W(j/m)$ is the triweight function. He concluded that, for very large m, the weight diagram is approximately $(315/512)*W(j/m)\ (3 - 11(j/m)^2)$ equivalent to the kernel given by Müller (1984).

Dagum and Bianconcini (2008) provided different kernel characterizations of the Henderson filter following the Reproducing Kernel Hilbert Space (RKHS) methodology.

A RKHS is a Hilbert space characterized by a kernel that reproduces, via an inner product, every function of the space or, equivalently, a Hilbert space of real-valued functions with the property that every point evaluation functional is bounded and linear. Berlinet and Thomas-Agnan (2003) showed how the space of polynomials of degree at most p is a RKHS. At this regard, the minimization problem (3) can be rewritten in continuous time as follows

$$\min_{g \in P_3} \|y - g\|^2_{P_3} = \int_T (y(t-s) - g(t-s))^2 f_0(s)\, ds \tag{7}$$

where $\|\cdot\|^2_{P_3}$ denotes the norm in the space of polynomials of degree at most 3, P_3, and f_0 is the density corresponding to the weighting function W_j.

Theorem 1. *The minimization problem (7) has a unique and explicit solution given by*

$$\hat{g}_t = \int_T y(t-s)^2 K_4(s)\, ds \tag{8}$$

where K_4 is the fourth (also third for the symmetry of f_0) order kernel representation of the Henderson filter.

The reader is referred to Dagum and Bianconcini (2008) for a detailed proof of Theorem 1. The kernel K_4 can be written as products of the reproducing kernel $R_3(t, \cdot)$ of the space $\mathbf{P}_3$ and the density function f_0, that has finite moments up to order 8. In particular, using the Christoffel-Darboux formula, for the sequence $(P_i)_{0 \le i < 3}$ of orthonormal polynomials with respect to the density f_0,

$$K_4(t) = \sum_{i=0}^{2} P_i(t)\, P_i(0)\, f_0(t) \tag{9}$$

The density corresponding to W_j and its orthonormal polynomials have to be determined. As shown by Bianconcini (2006), the former is given by

$$f_{0H}(t) = \frac{m+1}{k} W((m+1)\,t), \qquad t \in [-1,1] \tag{10}$$

here $k = \int_{-m-1}^{m+1} W(j)\, dj$, and $j = (m + 1)\, t$. On the other hand, the polynomials were computed by solving the Hankel system based on the moments of the density f_{0H} (Brezinski 1980).

The kernel K_4 provides the exact continuous representation for the Henderson filters, and, when $m = 6$, it results

$$K_4(t) = \frac{45}{7199963584}(-51619 - 156541t^2)(-1+t^2)(-64+49t^2)(-81+49t^2)$$

The main drawback of this formulation is that it depends on the length of the filter, and it needs to be calculated any time that m changes.

The asymptotic equivalent kernel representation based on the triweight density function allows to overcome such limitation. However, it often provides poor approximations for finite sample sizes (small values of m). Dagum and Bianconcini (2008) derived an equivalent reproducing kernel representation based on the biweight function $f_{0B}(t) = \frac{15}{16}(1-t^2)^2$, $t \in [-1,1]$, and showed that it closely approximates the Henderson filter when the length is small, such as up to 23 terms.

2.1. Choice of the Kernel Function

We compare the biweight and triweight functions with the exact one in order to evaluate the goodness of the approximation in terms of L_2-norm (Rice, 1983)

$$\|f-g\|_{L_2} = \sqrt{\int_{-1}^{1} (f(t)-g(t))^2\, dt}\; (\cdot)$$

These norms depend in a complicated manner on the length of the filter, and an evaluation for several values of *m* is provided in Table 1 for both the densities and third order kernels.

TABLE 1
L^2-norms for different values of m.

	m					
	4	**6**	**11**	**100**	$\to\infty$	$\to 0$
Densities						
Exact-Triweight	0.115	0.091	0.060	0.008	0.000	0.237
Exact-Biweight	0.013	0.032	0.063	0.114	0.122	0.119
3rd. order kernel						
Exact-Triweight	0.168	0.137	0.093	0.014	0.000	0.309
Exact-Biweight	0.022	0.026	0.066	0.145	0.159	0.156

It can be noticed that, for small values of *m*, the biweight provides a better approximation in terms of both density and third order kernel than the triweight ones. In particular, as shown in Figure 1, the distance between the exact and the biweight third order functions reaches its minimum when $m = 5.687$, whereas the distance with the triweight kernel approach to zero as $m \to \infty$, indicating that it represents the asymptotical equivalent kernel of the Henderson density and filter. Furthermore, it is clear from the figure, that the biweight kernel provides a better representation for Henderson kernels for *m* values equal to or less than 14, whereas beyond this value the triweight closer approximates the exact kernel. Hence, for small filter length, such as 9,13 and 23 terms, the biweight can be used to approximate the weight diagram of the Henderson, whereas for longer filters, the triweight will be a better choice.

2.2. Choice of the Bandwidth Parameter

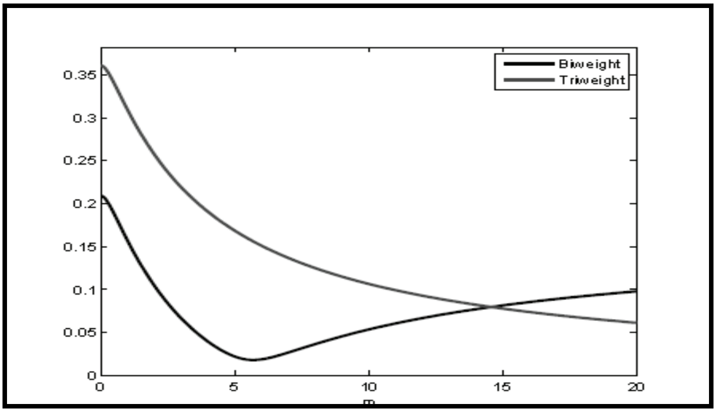
In kernel estimation, the weights of the linear filter are determined as follows

$$w_j = \frac{K(j/b)}{\sum_{j=-m}^{m} K(j/b)} \qquad j = -m, \ldots, m \tag{11}$$

where *K* is the kernel function, and *b* is the bandwidth parameter chosen in order to ensure a filter length equal to $2m + 1$.

FIGURE 1

L^2-norm between exact and triweight as well as biweight third order kernels.



The choice of the kernel function has been previously discussed. The order of the applied kernel depends on the degree of the fitted polynomial, and since the Henderson filter was build to be an unbiased estimator of a cubic polynomial, third order kernel representations are taken into account.

We now address inferential issues related to the choice of optimal bandwidth parameters b according to specific (unbiasedness and smoothing) properties of the filter weights. In practice, for linear estimators, this parameter is chosen a priori in order to ensure a prespecified filter length.

2.2.1. *Optimal Bandwidth*

Following the derivation of the exact density function, Dagum and Bianconcini (2008) found that when the support of the kernel function is $[-1,1]$, a bandwidth parameter equal to $(m + 1)$ represents a good choice, since it allows to cover at least the 95% of the area covered by the kernel functions. However, since the discrete approximation of the continuous kernel functions do not perfectly cover all the area under the curve, $\sum_{j=-m}^{m} j^2 w_j$ differs to zero, as shown in Table 2. In other words, the filter weights do not satisfy the property of reproducing a cubic polynomial trend without distortion as required by Henderson (1916).

TABLE 2

$\sum_{j=-m}^{m} j^2 w_j$ values of ideal and kernel weights for different values of m.

	m				
	4	**6**	**11**	**100**	$\to \infty$
Weights					
Biweight	0.050	0.026	0.009	0.000	0.000
Triweight	–0.019	–0.011	–0.004	0.000	0.000

In order to overcome such limitation, given a filter length equal to $2m + 1$, the bandwidth parameter could be selected as to optimize the Henderson problem

$$\min \sum_{j=-m}^{m} (\Delta^3 w_j)^2 \tag{12}$$

subject to

$$\sum_{j=-m}^{m} j^2 w_j = 0 \tag{13}$$

Table 3, and 4 provide an evaluation for different filter lengths and optimal bandwidths of eq. (12) and eq. (13) when the biweight and triweight third order kernels are applied, respectively.

TABLE 3
Properties of biweight weights based on optimal bandwidth for different values of *m*.

	m		
	4	**6**	**11**
$\sum_{j=-m}^{m} (\Delta^3 w_j)^2$	0.101	0.012	3.73e-4
$\sum_{j=-m}^{m} j^2 w_j$	0.000	0.000	0.000
Optimal bandwidth	4.927	6.951	11.973

TABLE 4
Properties of triweight based on optimal bandwidth for different values of *m*.

	m		
	4	**6**	**11**
$\sum_{j=-m}^{m} (\Delta^3 w_j)^2$	0.158	0.015	3.44e-4
$\sum_{j=-m}^{m} j^2 w_j$	0.000	0.000	0.000
Optimal bandwidth	5.102	7.122	12.139

It can be noticed that all the filters are now unbiased estimators of a cubic polynomial trend $\left(\sum_{j=-m}^{m} j^i w_j = 0 \quad i = 1,2,3\right)$, and also that the biweight presents smaller smoothness values than the triweight third order kernel for small filter lengths. On the other hand, in terms of optimal bandwidths, if we round the bandwidth parameters to the greatest integer value, we obtain exactly $(m + 1)$ for each kernel and filter length.

3. HENDERSON KERNELS AT THE BOUNDARIES

The derivation of the two-sided symmetric Henderson filter has assumed the availability of $2m + 1$ observations centered at t. Obviously, for a given finite sequence $\{y_t, t = 1, ..., T\}$, it is not possible to obtain the estimates of the signal for the (first and) last m time points, which is inconvenient since we are particularly interested at the most recent estimates. Several approaches have been followed in the literature to derive the asymmetric filters associated to the symmetric Henderson average.

As a local polynomial estimator, they have been obtained by fitting a local polynomial to the available observations $y_t, t = t - m + 1, \ldots, T,$ that is by minimizing

$$\sum_{j=-m}^{q} W_j (y_{t+j} - \beta_0 - \beta_1 j - \cdots - \beta_p j^p)^2, \qquad q = 0,...,m-1$$

where W_j denote the weighting function given in eq. (5). The corresponding weights satisfy the cubic polynomial reproduction constraints as in the interior of time support, but the preservation of these bias properties is done at the expenses of the variance, which is very high, since most of the contribution to the trend estimate comes from the current observation (see *i.e.* Proietti and Luati, 2008).

In the X11ARIMA seasonal adjustment procedure, Dagum (1982) proposed to apply the symmetric two-sided filter to the series extended by m forecasts. This is safer providing that we are capable of producing optimal forecasts according to some parametric or nonparametric model (*e.g.* by fitting a time series model of the ARIMA class). An intuitive and easily established fact is that if the forecasts are optimal in the mean square error sense, then the variance of the revision is a minimum (see Wallis, 1983).

A different and wider applied method can be found in Musgrave (1964). This technique is applied in the X-11 to derive trend-cycle averages corresponding to the Henderson filter. The approach assumes that, given $2m + 1$, the last data points of the time series follow a simple linear trend. Then the asymmetric weights $\{v_{-m}, ..., v_q\}$ are derived to minimize the expected square revision error

$$E\left(\sum_{j=-m}^{q} v_j y_j - \sum_{j=-m}^{m} w_j y_j \right)^2$$

under the constraint that $\sum_{j=-m}^{q} v_j = 1$. A complete and detailed review on this method can be found in Doherty (2001) and Quenneville et al. (2003). In particular, the latter showed that these filters are equivalent to apply the Henderson symmetric average on the available $m + q + 1$ data points with the missing last $m - q$ points replaced by forecasts obtained as linear combinations of the first available $m + q + 1$ data points.

3.1. Boundary Kernels

A different derivation of the asymmetric Henderson filters can be provided by taken into account for the reproducing kernel representation of the filter as introduced in the previous Section. At this regard, the Henderson filter is transformed into a continuous kernel function, and, at each time point t^*, the estimate $\hat{y}_{t^*}$ is obtained as a weighted mean of the sample $\{y_t, t = 1, ..., T\}$ with observations close to t^* receiving the largest weights. The latter are provided by the kernel function K defined on $[-1, 1]$, and specified only up to an unknown smoothing parameter b. That is (Nadaraya, 1964; Watson, 1964).

$$\hat{y}_t = \frac{\sum_{t=1}^{T} K\left(\frac{t-t^*}{b}\right) y_t}{\sum_{t=1}^{T} K\left(\frac{t-t^*}{b}\right)} \cong \frac{\sum_{j=-m}^{m} K\left(\frac{j}{b}\right) y_{t+j}}{\sum_{j=-m}^{m} K\left(\frac{j}{b}\right)} (\cdot) \tag{14}$$

The equivalence between the two formulations in eq. (14) derives from the fact that the bandwidth b is selected to ensure that only $2m + 1$ observations surrounding the target point will receive nonzero weights.

The approach is quite simple, but presents some flaws. Indeed, at the boundary of the predictor space $(t = 1, \dots, m,$ and $t = t - m + 1, \dots, T)$, the kernel neighborhood is asymmetric ($j = -m, ..., q$; $q = 0, ..., m - 1$) and the estimates may have substantially bias and may show an increase in the variance. Therefore, a variety of kernel modifications have been proposed in the literature to provide approximated and asymptotic adjustments to overcome these drawbacks.

In order to derive the asymmetric weights associated to the symmetric Henderson average, Dagum and Bianconcini (2006, 2008, and 2010) applied the "cut and normalize" method proposed by Gasser and Muller (1979). The latter showed that, for a fixed symmetric filter length $2m + 1$, and corresponding bandwidth b, the effective domain of the kernel function in the boundaries is $[-1, q^*)$, where $q^* = q/b$, instead of $[-1, 1]$ as for an interior point.

For simplicity, only the left boundary effects, *i.e.* $q^* < 1$, will be discussed here. The right boundary effects proceed in the same manner. Since $s = j/b \in [-1, q^*)$, the symmetry of the kernel is lost, such that $\int_{-1}^{q^*} K(s)\, ds \neq 1$ and $\int_{-1}^{q^*} s^i K(s)\, ds \neq 0$, $i = 1, \dots, p - 1$.

The boundary kernel K_{q^*} is obtained by "cutting" the symmetric kernel K to omit that part of the function lying between q^* and 1, and by "normalizing" it between -1 and q^* as follows

$$K_{q^*}(s) = \frac{K(s)}{\int_{-1}^{q^*} K(s)\, dt} (\cdot) \tag{15}$$

Boundary kernels (15) satisfy the following properties (Gasser and Muller, 1979).

1. **Moment conditions:** $\int_{-1}^{q^*} K_{q^*}(s)\,ds = 1, \int_{-1}^{q^*} K_{q^*}(s)\,s^i ds \neq 0, i = 1,\ldots,p.$

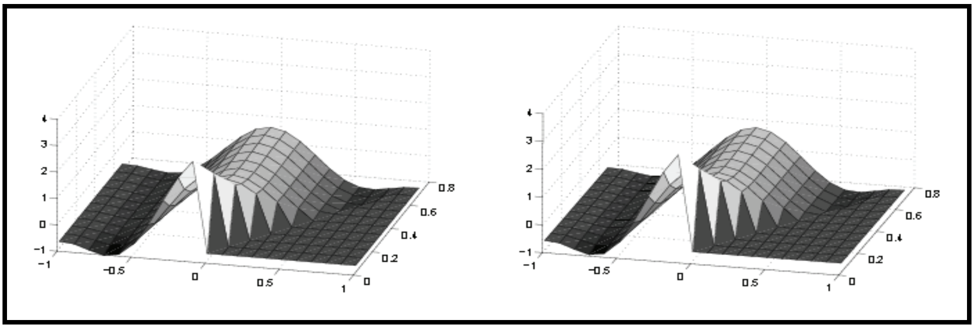
2. **Variance condition:** $\int_{-1}^{q^*} K_{q^*}^2(s)\,ds < c, \forall q^* \in [0,1),$ that is the asymptotic variance has to be uniformly bounded.

3. **Convergence:** the kernels depends continuously on q^*, and $K_{q^*} \xrightarrow[q^* \to 1]{} K.$

Property 3. is particularly appealing when asymmetric kernels are applied for trend-cycle estimation. In fact, a monotonic convergence to the corresponding symmetric filter should imply a reduction of the revisions when new observations are added to the series. On the other hand, Property 1. shows that these kernels are biased estimators of polynomial trend of degree *p* at the boundaries. However, since the neighborhood of observations is smaller than in the interior of the support, this can be allowed in order to reduce the variability of the estimates as shown by Property 2.

Boundary Henderson kernels can be derived starting from the exact, biweight and triweight third order kernel representations, as provided before. However, the former depends on the length of the filter, hence we now concentrate on the study of the boundary behavior of the biweight and triweight kernels. At this regard, Figure 2 shows their convergence to the corresponding symmetric kernel (q^* = 1). It can be noticed that, in both cases, the convergence is quite fast (in general after the previous to the last point asymmetric filter).

FIGURE 2
Biweight (left) and triweight (right) third order boundary kernels.



The application of the "cut and normalize" method brings to the following formula for the asymmetric weights

$$w_j = \frac{K(j/b)}{\sum_{j=-m}^{q} K(j/b)} \qquad j = -m,...,q; \quad q = 0,...,m-1 \tag{16}$$

where K is the symmetric kernel function, and b is the bandwidth parameter that can be different for each of the m asymmetric averages.

The choice of the kernel function is based on the properties of the corresponding boundary kernels. Biweight and the triweight represent good alternatives for deriving asymmetric Henderson filters for small filter lengths and asymptotically, respectively. The order of the kernels depends on the degree of the fitted polynomial in the interior points, and since the Henderson filter was built to be an unbiased estimator of a cubic polynomial in the interior points, third order kernels are taken into account.

3.2. Choice of the Bandwidth Parameter

The selection of the bandwidth parameter b is a crucial choice. In practice, for asymmetric filters, these parameters are chosen a priori in order to ensure a prespecified filter length $(m+q+1,\ q=0,...,m-1)$, as well as a rapid convergence to the corresponding symmetric filter.

3.2.1. *Fixed Bandwidth*

Dagum and Bianconcini (2006, 2008, and 2010) set b equal to m+1 for each of the m asymmetric filters. The properties of the resulting weights are compared in terms of Euclidean distances with the asymmetric filters derived via local polynomial regression (Direct Asymmetric Filters, DAF) and with the Musgrave ones. With reference to the last point asymmetric filter (q = 0), these measures are displayed in Table 5 and 6 for the DAF and Musgrave weights, respectively.

TABLE 5
Euclidean distances of the last point biweight and triweight weights and last point DAF for different values of *m*.

Kernels	***m***		
	4	**6**	**11**
Biweight	0.66	0.72	0.73
Triweight	0.62	0.68	0.70

TABLE 6
Euclidean distances of the last point biweight and triweight weights and last point Musgrave filter for different values of *m*.

Kernels	*m*		
	4	**6**	**11**
Biweight	0.13	0.07	0.10
Triweight	0.14	0.10	0.11

As we can notice, the boundary weights are much closer to the Musgrave asymmetric filter than to the DAF weights. This is clear by the fact that, while DAF filters are built in order to preserve the unbiasedness property as in the interior points, the Musgrave and boundary weights do not satisfy this condition. They try to find a better compromise between bias and variance in the boundaries. In particular, the RKHS weights inherit the same moment properties of the corresponding boundary kernels: independently on the filter length, they just pass a constant given the fact that they sum to one. On the other hand, given the asymmetry of the kernels in the boundaries, the unbiasedness properties are not satisfied, that is $\sum_{j=-m}^{q} j^i w_j \neq 0,\ i = 1,2,3.$ In this way, they are similar to the Musgrave filters that try to reproduce a linear trend in the boundaries by shrinking the slope toward zero, but some bias is present in the estimates.

Table 7 provides the values of $\sum_{j=-m}^{q} j^i w_j,\ i = 1,2,3,4.$ for the exact, biweight and triweight asymmetric weights associated to the 9-term symmetric filter.

TABLE 7
Properties of several asymmetric filters associated with the symmetric 9-term Henderson filter.

q	**0**	**1**	**2**	**3**	*q*	**0**	**1**	**2**	**3**
$\sum_{j=-m}^{q} jw_j$					$\sum_{j=-m}^{q} j^2w_j$				
Exact	–0.50	–0.08	0.17	0.15	Exact	0.13	0.37	0.79	0.74
Biweight	–0.48	–0.06	0.18	0.15	Biweight	0.04	0.31	0.74	0.65
Triweight	–0.40	0.005	0.18	0.09	Triweight	–0.01	0.28	0.60	0.34
$\sum_{j=-m}^{q} j^3w_j$					$\sum_{j=-m}^{q} j^4w_j$				
Exact	2.04	1.75	2.46	2.31	Exact	–11.99	–8.31	–5.53	–6.02
Biweight	2.35	1.97	2.66	2.41	Biweight	–13.08	–9.05	–6.18	–7.09
Triweight	1.92	1.66	2.20	1.45	Triweight	–9.82	–6.69	–4.74	–7.34

It can be noticed that all the filters just pass a constant and will be biased estimators of polynomial trends of any order. Even if this is a major drawback in the interior points of the sample period, this could be a good property at the end where just few observations are available for the estimation of the trend.

Furthermore, the bias introduced by the asymmetric filters can be compensated by a better performance in terms of variances of the estimates, which should be smaller with respect to the DAF, and similar to the Musgrave filters. For this reason, Table 8 shows the leverage values corresponding to last point biweight, triweight, DAF, and Musgrave asymmetric weights associated to symmetric filters of different lengths.

TABLE 8
Leverages of last point asymmetric filters associated to several symmetric Henderson filters.

Weights	*m*		
	4	**6**	**11**
DAF	0.99	0.96	0.83
Musgrave	0.58	0.42	0.29
Biweight	0.49	0.38	0.24
Triweight	0.54	0.42	0.27

As expected, the values are smaller for the RKHS weights than the DAF and close to the Musgrave last point filters. Furthermore, for any filter length, the biweight filter provides smaller leverages than the triweight one. To compare the theoretical properties of different smoothers the analysis of their behavior with respect to cyclical components is generally performed. The set of weights $\{w_j, j = -m, ..., q\}$ is called the *impulse response function* of the filter, and its properties are described in the frequency domain by looking at its Fourier transform, $H(\omega)$, called the *frequency response function*. That is,

$$\hat{y}_t = \sum_{j=-m}^{q} w_j y_{t+j} = \sum_{j=-m}^{q} w_j \exp(-i\omega(t+j)) = H(\omega)\exp(-i\omega t) = H(\omega)\, y_t$$

$H(\omega)$ fully describes the effect of the linear filter on the given input (Jenkins and Watts, 1968). In particular,

$$H(\omega) = G(\omega)\exp(-i\phi(\omega)) \tag{17}$$

where $G(\omega)$ is called the *gain function* of the filter and $\phi(\omega)$ is the *phase shift function*. $G(\omega)$ describes how much the amplitudes of a time series are amplified or reduced at the frequency ω, whereas $\varphi(\omega)$ contains information about the translation on time of the cyclical components.

In order to analyze and compare the properties of the boundary asymmetric kernels, we take into account for their frequency response function, that incorporates both the information coming from the gain function in terms of revisions, as well as the information coming from the phase shift function in terms of delay in detecting true turning points. Hence, we compute the Euclidean distance between the frequency response function of the asymmetric filter $H_a(\cdot)$ and the frequency response function of the corresponding symmetric Henderson filter $H(\cdot)$, that is

$$\sqrt{\int_0^{1/2} \left| H_a(\omega^*) - H(\omega^*) \right|^2 d\omega^*} \tag{18}$$

Table 9 illustrates the values of these distances for the biweight and triweight boundary weights associated to the 9-, 13-, and 23-term symmetric filters derived by the same kernels with bandwidth equal to $m + 1$.

TABLE 9
Euclidean distances between the frequency response function of the biweight and triweight boundary kernels and the frequency response function of the corresponding 9-term, 13-term and 23-term symmetric filter.

q	**0**	**1**	**2**	**3**							
Biweight	0.11	0.02	0.03	0.03							
Triweight	0.09	0.01	0.03	0.02							
q	**0**	**1**	**2**	**3**	**4**	**5**					
Biweight	0.18	0.08	0.02	0.04	0.04	0.02					
Triweight	0.15	0.06	0.02	0.04	0.03	0.01					
q	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Biweight	0.33	0.23	0.15	0.08	0.04	0.04	0.05	0.05	0.03	0.02	0.01
Triweight	0.30	0.20	0.12	0.05	0.03	0.04	0.05	0.04	0.02	0.01	0.01

The highest distances are observed for the last point asymmetric weights both for the biweight and the triweight boundary filters. However, these measures drastically reduce as q increases, indicating a fast convergence of the asymmetric filters to the corresponding symmetric ones.

3.2.2. *Optimal Bandwidth*

The asymmetric filters derived by imposing a bandwidth parameter equal to $(m + 1)$ present good properties in terms of convergence to the associated symmetric filters. However, in order to improve the convergence of the boundary filters, optimal bandwidths, specific for each asymmetric filter, can be selected in order to minimize eq. (18).

Table 10 illustrates the optimal values selected for each of the biweight and triweight boundary filters corresponding to the 9- and 13-term biweight and triweight symmetric weights, the latter computed by selecting the optimal bandwidths given in Table 3 and 4. We can notice that the bandwidths are quite close for the several asymmetric filters, with values ranging from m to $m + 1$.

TABLE 10
Optimal bandwidths for the biweight and triweight boundary kernels associated to 9-term and 13-term biweight and triweight optimal symmetric filter.

q	0	1	2	3		
Biweight	4.01	4.27	5.07	4.09		
Triweight	4.01	4.69	4.01	4.26		
q	0	1	2	3	4	5
Biweight	6.01	6.01	6.36	6.01	6.01	6.46
Triweight	6.01	6.01	6.77	6.01	6.01	6.64

Looking at the spectral properties of the related asymmetric weights, Table 11 shows that the convergence of the filters to the associated symmetric ones is improved mainly for the last point filter, and for $q \geq 1$ the distances are almost null.

TABLE 11
Euclidean distances between the frequency response function of the biweight and triweight boundary kernels and the frequency response function of the corresponding 9-term and 13-term symmetric filters.

q	0	1	2	3		
Biweight	0.06	0.01	0.02	0.00		
Triweight	0.07	0.01	0.03	0.00		
q	0	1	2	3	4	5
Biweight	0.12	0.03	0.02	0.03	0.01	0.00
Triweight	0.14	0.05	0.01	0.03	0.02	0.00

As for the filters derived by fixing $b = m + 1$, the asymmetric weights based on optimal bandwidths do not satisfy the unbiasedness conditions as shown in Table 12 for those associated to a 9-term symmetric average. However, in this case, we can notice that these filters present values of the first moment smaller than before, and similar to the Musgrave filters. Hence, when optimal bandwidths are selected, the corresponding weights could pass a linear trend at the end of the sample period with small bias, mainly for the first and second asymmetric filter. Similar conclusions can be drawn for asymmetric weights associated to symmetric filter of any length, not reported here for space reasons.

On the other hand, the leverage values of these filters are higher than in the previous case when $b = m + 1$. In particular, for the 9-term symmetric filter, the weight associated to the current observation for the last point asymmetric filter is equal to 0.65 for both the biweight and triweight kernels. On the other hand, for the 13-term filter, this weight is equal to 0.48 and 0.50 for the biweight and triweight, respectively.

TABLE 12
Properties of optimal asymmetric filters associated with the symmetric 9-term Henderson filter.

q	0	1	2	3	q	0	1	2	3
$\sum_{j=-m}^{q} jw_j$					$\sum_{j=-m}^{q} j^2 w_j$				
Biweight	–0.35	0.02	0.17	0.00	Biweight	0.06	0.32	0.77	0.13
Triweight	–0.28	0.03	0.12	0.01	Triweight	0.01	0.28	0.34	0.05
$\sum_{j=-m}^{q} j^3 w_j$					$\sum_{j=-m}^{q} j^4 w_j$				
Biweight	0.98	1.27	2.61	0.08	Biweight	–4.44	–4.50	–6.03	–6.14
Triweight	0.94	1.40	1.07	0.10	Triweight	–3.81	–5.07	2.13	–5.90

TABLE 13
Euclidean distances of the optimal last point biweight and triweight weights and last point Musgrave filter for different values of *m*.

Kernels	***m***	
	4	**6**
Biweight	0.19	0.13
Triweight	0.23	0.17

In order to better evaluate the relationship of these optimal asymmetric weights with the Musgrave filters, we compute their Euclidean distances as shown in Table 13. We notice that these measures increase with respect to the case in which $b = m + 1$, since the leverage values of both the biweight and triweight last point filters increase. However, the spectral properties of these filters are improved with respect to the filters derived by imposing a fixed bandwidth, since they provide a better compromise in terms of both revisions (gain function) and detection of true turning points (phase shift function).

4. CONCLUDING REMARKS

The paper has considered the problem of estimating the trend of a time series in real time by means of reproducing kernel filters associated to symmetric Henderson averages. We showed that these filters share similar properties with the Musgrave surrogates adopted by X11 based seasonal adjustment procedures, that are known to minimize revisions for a certain class of time series. However, they are derived following a different optimization criteria with respect to the symmetric Henderson filter, with the consequence that the asymmetric filters do not converge monotonically to the symmetric one. At this regard, Dagum and Bianconcini (2006 and 2008) provided a different characterization of the Henderson weights within the Reproducing Kernel Hilbert Space (RKHS) methodology. According to this approach, a continuous kernel representation of the Henderson filter is obtained, and the same kernel function is used to derive both symmetric and asymmetric weights. The density function (*i.e.* a second order kernel) embedded on the linear filter is firstly determined. This is the starting point for obtaining higher order kernels, which are based on the product of the density and its orthonormal polynomials. For each kernel order, the asymmetric filters can be derived coherently with the corresponding symmetric weights.

Inferential issues have been addressed in terms of selection of both kernel order and bandwidth parameter. We have shown that the Henderson weight diagram can be generated using three different density functions: the exact related to the weighting function that appears in the Henderson minimization problem, the biweight and triweight densities. In particular, since the exact needs to be computed any time that the filter length changes, we suggest the use of the biweight for small filter lengths (less than 27 terms), and the triweight for longer filters. Dagum and Bianconcini (2008) also suggested to select a bandwidth parameter equal to m+1, and we have shown that this is almost equivalent to select b in order to minimize the sum of squares of the third differences under cubic polynomial reproduction constraints.

The asymmetric filters are derived by applying the same kernel functions adapted to the length of the filter. This approach has been introduced by Gasser and Muller (1979) (called "cut-and-normalized" method) to improve the properties of kernel estimators in the boundaries. We have shown that the corresponding asymmetric filters share similar properties to the Musgrave one in terms of polynomial reproduction. In particular, when the bandwidth parameters are all fixed to m+1, the former just pass a constant, whereas the latter a linear trend with small bias. On the other hand, when the filter-specific bandwidth parameters are selected in order to optimize the spectral properties of the asymmetric filters, most of the reproducing kernel filters also pass a linear trend with small bias.

Analyzing the frequency response functions of the asymmetric filters, the spectral properties of those obtained by means of reproducing kernels are better than those of filters obtained by local polynomial regression, and similar to the Musgrave ones.

REFERENCIAS BIBLIOGRÁFICAS

BERLINET, A. and THOMAS-AGNAN, C. (2003): *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers.

BIANCONCINI, S. (2006): *Trend-Cycle Estimation in Reproducing Kernel Hilbert Spaces*, Ph.D. Thesis, Department of Statistics, University of Bologna.

BREZINSKI, C. (1980): *Pade Approximation and General Orthogonal Polynomials*, Birkhauser, Basel.

DAGUM, E.B. (1982): "The effects of asymmetric filters of seasonal factor revisions", *Journal of the American Statistical Association*, 77, 732-738.

DAGUM, E.B. (1988): *The X11ARIMA/88 Seasonal Adjustment Method - Foundation and User's Manual*, Research Paper, Time Series Research and Analysis Division, Statistics Canada, Ottawa.

DAGUM, E.B. (1996): "A New Method to Reduce Unwanted Ripples and Revisions in Trend-Cycle Estimates from X11ARIMA," *Survey Methodology*, 22, 77-83.

DAGUM, E.B. and BIANCONCINI, S. (2006): Local polynomial trend-cycle predictors in reproducing kernel Hilbert spaces for current economic analysis, *Anales de Economia Aplicada*, Delta publicaciones, Madrid, pp. 1-22.

DAGUM, E.B. and BIANCONCINI, S. (2008): "The Henderson Smoother in Reproducing Kernel Hilbert Space", *Journal of Business and Economic Statistics*, 26 (4), 536-545.

DAGUM, E.B. and BIANCONCINI, S. (2010): *A unified probabilistic view of nonparametric predictors via reproducing kernel Hilbert spaces*, under review.

DOHERTY, M. (2001): "Surrogate Henderson filers in X-11", *Australian and New Zealand Journal of Statistics*, 43, 385-392.

FAN, J. and GIJBELS, I. (1996): *Local polynomial modelling and its applications*, Monographs on statistics and applied probability 66, Chapman and Hall/CRC.

FINDLEY, D.; MONSELL, B.; BELL, W.; OTTO, M. and CHEN, B. (1998): " New Capabilities and Methods of the X12ARIMA Seasonal Adjustment Program," *Journal of Business and Economic Statistics*, 16, pp. 127-152.

GASSER, T. and MULLER, H.G. (1979): "Kernel Estimation of Regression Functions". In *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757, eds. T. Gasser and M. Rosenblatt, Heidelberg: Springer-Verlag, pp. 23-68.

GRAY, A. and THOMSON, P. (1996): "Design of Moving-Average Trend Filters Using Fidelity and Smoothness Criteria," in *Time Series Analysis* (in memory of E.J. Hannan), eds. Robinson, P.M. and Rosenblatt, M., vol.II, New York: Springer Lecture Notes in Statistics, 115, 205-219.

HENDERSON, R. (1916): "Note on Graduation by Adjusted Average," *Transaction of Actuarial Society of America*, 17, 43-48.

KENDALL, M.G.; STUART, A. and ORD, J. (1983): *The Advanced Theory of Statistics*, Vol. 3, Ed. C. Griffin.

KENNY, P. and DURBIN (1982): "Local Trend Estimation and Seasonal Adjustment of Economic and Social Time Series," *Journal of the Royal Statistical Society A*, 145, 1-41.

LOADER, C. (1999): *Local Regression and Likelihood*, New York: Springer.

MÜLLER, H.G. (1984): "Smooth Optimum Kernel Estimators of Regression Curves, Densities and Modes," *Annals of Statistics*, 12, 766-774.

MUSGRAVE, J. (1964): "A *set of end weights to end all end weights*", Working paper. Washington DC: Bureau of Census.

NADARAYA, E.A. (1964): "On Estimating Regression". *Theory of Probability and its Applications*, 9 (1): 141-142.

PROIETTI, T. and LUATI, A. (2008): "Real time estimation in local polynomial regression, with application to trend-cycle analysis", *The Annals of Applied Statistics*, 2 (4), 1523-1553.

QUENNEVILLE, B. and LADIRAY, D. (2000): "*Locally adaptive trend-cycle estimation for X11*", Proocedings of the International Conference on Establishment Surveys-II, Buffalo, NY, June 2000.

QUENNEVILLE, B.; LADIRAY, D. and LEFRANCOIS, B. (2003): "A note on Musgrave asymmetrical trend-cycle filters", *International Journal of Forecasting*, 19, 727-734.

RICE, J. (1983): *Numerical Methods*, Software and Analysis, New York: McGraw-Hill.

SHISKIN, J.; YOUNG, A. and MUSGRAVE, J. (1967): "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper 15, US Department of Commerce, Bureau of the Census, Washington DC.

WALLIS, K. (1983): ”Models for X-11 and X-11 forecast procedures for preliminary and revised seasonal adjustments”, in *Applied Time Series Analysis of Economic Data*, A. Zellner ed., Washington DC: Bureau of the Census, 3-11, 1346-1370.

WATSON, G. (1964): ”Smooth regression analysis”. *Sankhya Series, A*, 26, 359-372.