

Local Polynomial Regression in Real Time ¹

Stima Concorrente nella Regressione Polinomiale Locale

Tommaso Proietti ²
S.E.F. e ME. Q.,

Università di Roma “Tor Vergata”, e-mail: tommaso.proietti@uniroma2.it

Alessandra Luati ³
Dipartimento di Scienze Statistiche “Paolo Fortunati”,
Università di Bologna, e-mail: luati@stat.unibo.it

Riassunto: Il lavoro concerne la stima concorrente nella regressione polinomiale locale. Si dimostra che l’adattamento automatico di filtri polinomiali agli estremi produce stime caratterizzate da elevata variabilità. La ragione risiede nel fatto che il filtro per la stima concorrente risulta fortemente localizzato. In alternativa, viene proposta una classe di filtri asimmetrici basati su un criterio di minimo errore quadratico medio di revisione soggetto a vincoli di riproduzione polinomiale. Tale classe approssima un dato filtro simmetrico, dipende da caratteristiche non note quali la pendenza e la curvatura del segnale, e comprende come casi particolari i filtri asimmetrici polinomiali e, quando lo stimatore simmetrico è il filtro di Henderson, i filtri surrogati di Musgrave. Un’analisi empirica evidenzia l’efficacia dei filtri asimmetrici appartenenti alla classe proposta.

Keywords: Henderson filter. Trend estimation. Fixed bandwidth. Musgrave asymmetric filters.

1. Introduction

The concern of the paper is real time estimation of the underlying trend in a time series by means of filters that arise from fitting a local polynomial of a given degree with a constant bandwidth. Real time estimation is of outmost importance in fields like economics and deals with the estimation of a signal at time t using the observations available up to and including time t . A well known property of local polynomial estimators is the automatic adaptation at the boundaries. It essentially means that the bias near the boundary is of the same order of magnitude as in the interior. It turns out, however, that for a local cubic fit, such as that arising from the well known filter due to Henderson (1916), the variance inflation resulting from the one-sided real time direct filter is very high, due to the fact that the filter is strongly localised at the current observations, with a leverage that is close to unity. The paper documents this basic feature and will be concerned in particular with the evaluation of alternative strategies aiming at the adaptation at the boundary of a given two-sided symmetric local polynomial filter. Our discussion will mostly refer to the Henderson filter. The latter has a long tradition for trend-cycle estimation in economic time series and still nowadays is employed for trend estimation in the X-11 cascade filter,

¹Financing from MIUR, Prot. 2004130572-2004, is gratefully acknowledged

²Address of correspondence: via Columbia 2, 00133 Roma

³Address of correspondence: via Belle Arti 41, 40126 Bologna

and as such it is an integral part of the X-12-ARIMA procedure, the official seasonal adjustment procedure in the U.S., Canada, the U.K. and many other countries.

The plan of the paper is the following. After reviewing the constructive principles presiding the derivation of the two-sided symmetric local polynomial filters, we provide a thorough assessment of the properties of the asymmetric filters automatically adapted at the boundary, which result from fitting a local polynomial with a fixed bandwidth to the observations available at the current time. The direct asymmetric filters can be equivalently derived through the reproducing kernel Hilbert space method, as we prove based on the Hankel representation of a reproducing kernel in the context of weighted least squares estimation. The key result, as we stressed above, is that the real time filter behaves differently from the other automatically adapted asymmetric filters inside the boundary. Section 3. evaluates an alternative class of fixed bandwidth asymmetric filters that result from minimising the mean square revision error subject to polynomial reproduction constraints. This class generalises the well known Musgrave's asymmetric approximation of the Henderson filters (Musgrave, 1964, see also Doherty, 2001), which is implemented in the seasonal adjustment filter X-11, developed by the U.S. Census Bureau. The class of filters depends on the properties of the true underlying signal, namely its level, slope, curvature, and so forth, which can be estimated from the data. In section 4. we provide a few illustrations dealing with economic time series. They address the issue of approximating the Henderson filter in real time and show that the slope and curvature play a relevant role for the derivation of the optimal real time approximation. These features can be estimated from the available data. The two features are on the contrary neglected by the Musgrave's asymmetric filters, which postulate that the true underlying signal is linear but only require that the approximate filter is capable of reproducing a zero degree polynomial. In section 5. we draw our conclusions.

2. Local polynomial filters and the Henderson Filter

Let us assume that time is discrete and that the series can be decomposed as $y_t = \mu_t + \epsilon_t$, where μ_t is the signal (trend), a smooth function of t , and $\epsilon_t \sim \text{NID}(0, \sigma^2)$ is the noise. The signal can be approximated locally by a polynomial of degree d in the distance j on a neighbourhood of time t , so that

$$y_{t+j} = m_{t+j} + \epsilon_{t+j}, \quad m_{t+j} = \beta_0 + \beta_1 j + \beta_2 j^2 + \cdots + \beta_d j^d, \quad j = 0, \pm 1, \dots, \pm h.$$

In matrix notation, the local polynomial approximation can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{N}(\mathbf{0}, \sigma^2 \mathbf{I}),$$

where $\mathbf{y} = [y_{t-h}, \dots, y_t, \dots, y_{t+h}]'$, $\boldsymbol{\epsilon} = [\epsilon_{t-h}, \dots, \epsilon_t, \dots, \epsilon_{t+h}]'$, the $r+1$ -th column of $\mathbf{X}$ is the $[(-h)^r, \dots, (-1)^r, \dots, (h-1)^r, (h)^r]'$, $r = 0, \dots, d$, and $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_d]'$. Setting $\mathbf{K} = \text{diag}(\kappa_{-h}, \dots, \kappa_{-1}, \kappa_0, \kappa_1, \dots, \kappa_h)$, the weighted least squares (WLS) estimate of the coefficients is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{y}$. In order to obtain $\hat{m}_t = \hat{\beta}_0$, we need to select the first element of the vector $\hat{\boldsymbol{\beta}}$. Hence, denoting by $\mathbf{e}_1$ the $d+1$ vector $\mathbf{e}_1' = [1, 0, \dots, 0]$, $\hat{m}_t = \mathbf{e}_1' \hat{\boldsymbol{\beta}} = \mathbf{e}_1' (\mathbf{X}'\mathbf{K}\mathbf{X})^{-1} \mathbf{X}'\mathbf{K}\mathbf{y} = \mathbf{w}'\mathbf{y} = \sum_{j=-h}^h w_j y_{t-j}$, which expresses the trend estimate as a linear combination of the observations with coefficients

$$\mathbf{w} = \mathbf{K}\mathbf{X}(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{e}_1.$$

The linear combination yielding the trend estimate is the local polynomial two-sided *filter*. It satisfies $\mathbf{X}'\mathbf{w} = \mathbf{e}_1$, or equivalently, $\sum_{j=-h}^h \mathbf{w}_j = 1$, $\sum_{j=-h}^h j^r \mathbf{w}_j = 0$, $r = 1, 2, \dots, d$. As a consequence, the filter $\mathbf{w}$ is said to preserve a deterministic polynomial of order d . Moreover, the filter weights are symmetric ($\mathbf{w}_j = \mathbf{w}_{-j}$), which follows from the symmetry of the kernel weights κ_j , and the assumption that the observations are equally spaced. The Henderson (1916) filter arises when $d = 3$ and $\kappa_j = [(h+1)^2 - j^2][(h+2)^2 - j^2][(h+3)^2 - j^2]$. The resulting weights $\mathbf{w}_j = \kappa_j(S_4 - S_2 j^2)(S_0 S_4 - S_2^2)^{-1}$, $j = 0, \pm 1, \dots, \pm h$, where $S_r = \sum_{j=-h}^h \kappa_j j^r$, maximise the smoothness of the estimated trend, in the sense that the variance of its third differences is minimum. It can be shown that the κ_j minimise the sum of squared third order differences of the weights sequence, $\mathbf{w}_j$.

2.1 Asymmetric filters and their automatic adaptation at boundary points

The derivation of the two-sided symmetric filter has assumed the availability of $2h + 1$ observations centred at t . Obviously, for a given finite sequence $y_t, t = 1, \dots, n$, it is not possible to obtain the estimates of the signal for the (first and) last h time points, which is inconvenient, since we are typically most interested at the most recent estimates.

Let us partition the matrices $\mathbf{X}$, $\mathbf{K}$ and the vector $\mathbf{y}$ as follows: $\mathbf{X} = [\mathbf{X}_p' \ \mathbf{X}_f']'$, $\mathbf{y} = [\mathbf{y}_p' \ \mathbf{y}_f']'$, $\mathbf{K} = \text{diag}\{\mathbf{K}_p; \ \mathbf{K}_f\}$, where $\mathbf{y}_p$ denotes the set of past available observations, whereas $\mathbf{y}_f$ is missing and $\mathbf{X}$ and $\mathbf{K}$ are partitioned accordingly. The direct asymmetric filter (DAF) arising as the solution to the above WLS problem is, in matrix notation,

$$\mathbf{w}_a = \mathbf{K}_p \mathbf{X}_p (\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)^{-1} \mathbf{e}_1.$$

The filter resulting from the automatic adaptation of the local polynomial fit can be equivalently derived using applying the symmetric filter $\mathbf{w}$ to the series extended by h forecasts, assuming that the future observations are generated according to a polynomial function of time of degree d , so that the optimal forecasts are generated by the same model. Under the local polynomial model the forecasted values of $\mathbf{y}_f$ is

$$\hat{\mathbf{y}}_f = \mathbf{X}_f (\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)^{-1} \mathbf{X}_p' \mathbf{K}_p \mathbf{y}_p.$$

Applying the two-sided filter $\mathbf{w}$ to the observations extended by the forecasts $\mathbf{y} = [\mathbf{y}_p' \ \hat{\mathbf{y}}_f']'$ where $\hat{\mathbf{y}}_f$ and using $(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1} = (\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p + \mathbf{X}_f' \mathbf{K}_f \mathbf{X}_f)^{-1} = (\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)^{-1} [\mathbf{I} + \mathbf{X}_f' \mathbf{K}_f \mathbf{X}_f (\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)^{-1}]^{-1}$ gives

$$\hat{m}_{t|t+q} = \mathbf{e}_1' (\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)^{-1} \mathbf{X}_p' \mathbf{K}_p \mathbf{y}_p,$$

which is also the estimate of the intercept of the polynomial that uses only the available information.

As an alternative, the direct asymmetric filters can be derived with the reproducing kernel Hilbert space (RKHS) approach (see Berlinet and Thomas-Agnan, 2004). In that context, the equivalent kernel of a linear estimator of order d can be obtained as $K_d(t) = R_d(t, 0) f_0(t)$, where $R_d(t, 0)$ is the reproducing kernel of a Hilbert space of polynomials up to degree $d \geq 1$ with inner product defined with respect to a density function $f_0(t)$. The reproducing kernel is so called because it reproduces any function in the Hilbert space in the sense that $\langle g, R_d(t, \cdot) \rangle_{\mathcal{H}} = g(t), \forall t \in T, g \in \mathcal{H}$, from which many inferential properties can be derived. Once $f_0(t)$ is chosen with finite moments $\nu_0, \nu_1, \dots, \nu_{2d}$, one way

to obtain the associated reproducing kernel is by means of Hankel determinants (Berlinet and Thomas-Agnan, 2004, Theorem 80) in that

$$K_d(t) = \frac{\det(\mathbf{H}_d^0[1, \mathbf{x}_t])}{\det(\mathbf{H}_d^0)} f_0(t)$$

where $\mathbf{H}_d^0$ is the Hankel matrix whose elements are the moments of $f_0(t)$, from ν_0 to ν_d in the first row and from ν_d to ν_{2d} in the last column, and $\mathbf{H}_d^0[1, \mathbf{x}_t]$ is the matrix obtained replacing the first column of $\mathbf{H}_d^0$ by the vector $\mathbf{x}_t = [1, t, t^2, \dots, t^d]'$. In our discrete setting, choosing the (normalised) Henderson kernel κ_j in place of the density $f_0(t)$, then $\nu_r = S_{qr}$ for $r = 0, \dots, 2d$ and the matrix $\mathbf{H}_d^0$ becomes $\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p$, so that the elements of the filter $\mathbf{w}_a$, for $j = -h, \dots, q$ and $q = 0, \dots, h$ are given by

$$\mathbf{w}_{a,j} = \frac{\det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p[1, \mathbf{x}_j])}{\det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)} \kappa_j$$

where $\mathbf{x}_j = [1, j, j^2, \dots, j^d]'$. The above expression is exactly the same that we would obtain by solving for $\hat{\beta}_0 = \hat{m}_t$ the least squares equation $(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p) \hat{\beta} = \mathbf{X}_p' \mathbf{K}_p \mathbf{y}_p$ using the Cramer rule for the explicit solution of a linear system. In fact, setting $\mathbf{b} = \mathbf{X}_p' \mathbf{K}_p \mathbf{y}_p$, the first coordinate of the solution vector is $\hat{\beta}_0 = \det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p[1, \mathbf{b}]) \det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)^{-1}$. As $\mathbf{b} = \sum_{j=-h}^q \mathbf{x}_j \kappa_j y_{t+j}$, then $\det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p[1, \mathbf{b}]) = \sum_{j=-h}^q \det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p[1, \mathbf{x}_j]) \kappa_j y_{t+j}$ and therefore

$$\hat{m}_t = \sum_{j=-h}^q \frac{\det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p[1, \mathbf{x}_j])}{\det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)} \kappa_j y_{t+j}.$$

This result also holds for symmetric filters, arising when $q = h$, and for any choice of the kernel κ_j , providing an alternative way to express both the trend estimate and the equivalent kernel of the linear filter resulting by weighted linear regression.

2.2 Properties of the direct asymmetric filters

Partitioning the weights of the two-sided symmetric filter in two groups, $\mathbf{w} = [\mathbf{w}_p', \mathbf{w}_f']'$, where $\mathbf{w}_p$ contains the weights attributed to the past and current observations and $\mathbf{w}_f$ those attached to the future unavailable observations, we obtain the fundamental relationship which states how the asymmetric filter weights are obtained from the symmetric ones

$$\mathbf{w}_a = \mathbf{w}_p + \mathbf{K}_p \mathbf{X}_p (\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)^{-1} \mathbf{X}_p' \mathbf{w}_f.$$

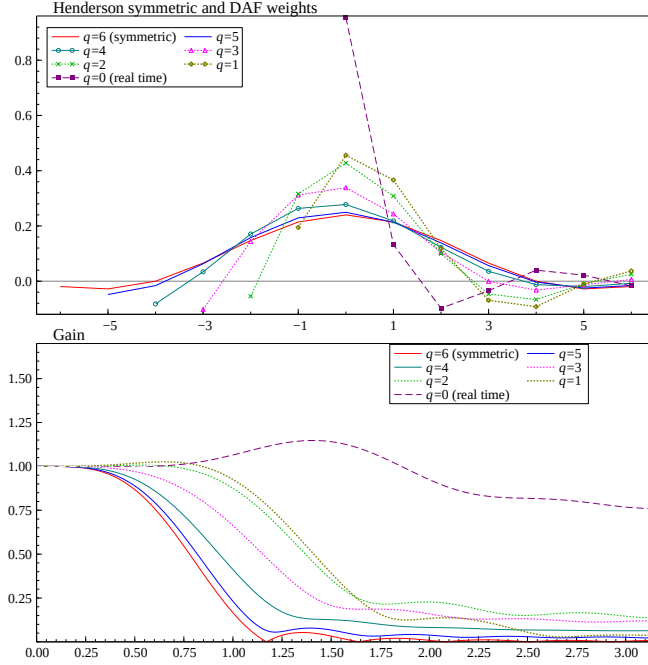
Premultiplying both sides by $\mathbf{X}_p'$, we can see that the asymmetric filter weights satisfy the following polynomial reproduction constraints: $\mathbf{X}_p' \mathbf{w}_a = \mathbf{X}_p' \mathbf{w}_p + \mathbf{X}_p' \mathbf{w}_f = \mathbf{X}' \mathbf{w}$. If the design of the time points is centred around the current time then $\mathbf{X}' \mathbf{w} = \mathbf{e}_1$. Thus, the bias in estimating an unknown function of time has the same order of magnitude as in the interior of time support.

We now show that the weights $\mathbf{w}_a$ are the unique minimisers with respect to $\mathbf{v}$ of the following constrained problem:

$$\min_{\mathbf{v}} (\mathbf{v} - \mathbf{w}_p)' \mathbf{K}_p^{-1} (\mathbf{v} - \mathbf{w}_p) \quad \text{s.t. } \mathbf{X}_p' \mathbf{v} = \mathbf{X}' \mathbf{w},$$

where $\mathbf{w} = [\mathbf{w}_p', \mathbf{w}_f']'$. The first order conditions give $\mathbf{v} = \mathbf{w}_p + \mathbf{K}_p \mathbf{X}_p \mathbf{l}$, where $\mathbf{l}$ is a vector of Lagrange multipliers. Premultiplying both sides by $\mathbf{X}_p'$ and replacing $\mathbf{X}_p' \mathbf{v} =$

Figure 1: Gain, phase and weights for the symmetric and asymmetric Henderson filters w_a ; q is the number of future observations available for estimating the signal.



$\mathbf{X}'\mathbf{w}$ gives $\mathbf{l} = (\mathbf{X}_p'\mathbf{K}_p\mathbf{X}_p)^{-1}\mathbf{X}_f'\mathbf{w}_f$, and replacing into the expression for $\mathbf{v}$ gives $\mathbf{v} = \mathbf{w}_a$. Hence, the asymmetric weights w_a minimise the weighted distance between the asymmetric filter coefficients and the symmetric ones, where the weights are provided by the reciprocal of the kernel weights. This minimum distance property is useful in order to compare the DAF with the class of asymmetric filters derived in section 3.

Figure 1 plots the weights of the direct asymmetric Henderson filter for q ranging from 0 (real time filter) to h (symmetric Henderson filter), along with their gain when the bandwidth takes the value $h = 6$, producing the Henderson 13 terms moving average when all the necessary observations are available. The real time filter uses 7 consecutive observations and it is very much concentrated on the current observation. As a consequence the gain behaves rather poorly, being close to one also at the high frequencies.

Hence, our analysis reveals that there is a sort of discontinuity in the behaviour of the filter, when we move from $q = 0$ (real time filter) to $q = 1$ (one future observation is available). The real time filter is unbiased if the series is generated by a cubic polynomial; however, the preservation of the 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. This can be explained by means of the following relation, that gives the leverage of the filter, i.e. the weight attached to the observation to be estimated,

$$w_{a,0} = \kappa_0 \mathbf{e}_1' (\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)^{-1} \mathbf{e}_1 = \kappa_0 \frac{\det(\mathbf{M}_{1,1})}{\det(\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p)}$$

where $\mathbf{M}_{1,1}$ is the submatrix obtained by deleting the first row and column of $\mathbf{X}_p' \mathbf{K}_p \mathbf{X}_p$.

(i) For fixed values of d , the leverage decreases as long as the span of the filter increases. It is maximum for the real time filter ($q = 0$) and minimum for the symmetric filter ($q = h$).

(ii) On the other hand, for fixed values of h or q the leverage increases if the degree of the fitting polynomial increases. It is minimum for $d = 0$ and maximum for $d = h$.

In particular, $w_{a,0} = 1$ for $d = h$. The latter equality can be proved by noticing the general fact that the $h + 1$ -th row of $\mathbf{X}_p$ (last row when real time filters are considered), whose elements correspond to j^r , $r = 0, \dots, d$, is the vector $\mathbf{e}'_1 = [1, 0, 0, \dots, 0]$. Given that $\mathbf{K}_p$ is diagonal, it follows from the row-column matrix product that $\mathbf{M}_{1,1} = \mathbf{X}_{h+1,1}' \mathbf{K}_{h+1,h+1} \mathbf{X}_{h+1,1}$ where $\mathbf{X}_{i,j}$ and $\mathbf{K}_{i,j}$ are submatrices obtained by deleting the i -th row and j -th column of $\mathbf{X}_p$ and $\mathbf{K}_p$. If $d = h$, then $\mathbf{X}_p$ and $\mathbf{X}_{h+1,1}$ are square matrices that have different dimensions but same determinant, as it is immediate to see by calculating $\det(\mathbf{X}_p)$ from the last row of $\mathbf{X}_p$ with the Laplace formula. Hence, it follows from the Binet-Cauchy theorem that

$$\frac{\det(\mathbf{M}_{1,1})}{\det(\mathbf{X}'_p \mathbf{K}_p \mathbf{X}_p)} = \frac{\det(\mathbf{X}_{h+1,1}') \det(\mathbf{K}_{h+1,h+1}) \det(\mathbf{X}_{h+1,1})}{\det(\mathbf{X}'_p) \det(\mathbf{K}_p) \det(\mathbf{X}_p)} = \frac{1}{\kappa_0}$$

and therefore $w_{a,0} = 1$. Since the filter reproduces polynomials up to the order d , $w_{a,j} = 0$ for $j = -h, \dots, -1$. This result holds even for symmetric and nearest neighbour filters, where the maximum value d can assume is $2h$. The proof of (i) and (ii) is based on a generalised version of the Binet-Cauchy theorem, that consents of calculating the determinant of the product of rectangular matrices,

$$\frac{\det(\mathbf{M}_{1,1})}{\det(\mathbf{X}'_p \mathbf{K}_p \mathbf{X}_p)} = \frac{\sum_{\pi_j}^{\binom{h+q}{d}} \det(\mathbf{X}_{h+1,1,\pi_j}') \det(\mathbf{K}_{h+1,h+1,\pi_j}) \det(\mathbf{X}_{h+1,1,\pi_j})}{\sum_{\pi_j}^{\binom{h+q+1}{d+1}} \det(\mathbf{X}'_{\pi_j}) \det(\mathbf{K}_{\pi_j}) \det(\mathbf{X}_{\pi_j})}$$

where $\mathbf{X}'_{\pi_j}$ denotes a square submatrix of $\mathbf{X}'_p$ obtained taking all its rows and $d + 1$ columns chosen on the set π_j of the $h + q + 1$ columns of $\mathbf{X}'_p$ and the summation $\sum_{\pi_j}^{\binom{h+q+1}{d+1}}$ is extended to the $\binom{h+q+1}{d+1}$ subsets of $1, \dots, h + q + 1$ with $d + 1$ elements; $\mathbf{K}_{\pi_j}$ is the square submatrix of $\mathbf{K}_p$ whose $d + 1$ columns (and rows) correspond to those chosen for $\mathbf{X}'_{\pi_j}$.

3. On a general class of asymmetric filters

We now consider a class of asymmetric filters approximating a given symmetric two-sided smoothing filter. The class depends on unknown features of the series, such as slope and curvature, which can be estimated from the data, and encompasses the Musgrave's surrogate filters. The latter approximate the two-sided Henderson filter at the end of the sample and are a component of the well-known X-11 cascade seasonal adjustment filter. The minimum mean square revision error strategy which is at the basis of our criterion was originally proposed by Musgrave (1964). Gray and Thomson (2002) generalised this idea to the case of a series generated by a local dynamic model. We propose a different derivation of Gray and Thomson result that is more general and clarifies some issues of the design of asymmetric filters, among which the connections with the DAF. We also provide an alternative expression for the asymmetric weights that is directly connected to Musgrave's result. Furthermore, our result is readily generalisable to the nearest neighbour bandwidth case.

Assume that the observations are generated as $\mathbf{y} = \mathbf{U}\boldsymbol{\gamma} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{D})$, where $\mathbf{U}$, $\mathbf{Z}$ are suitable design matrix. We aim at determining the asymmetric filter $\mathbf{v}$ minimising

the mean square revision error subject to constraints. The constraints are specified as follows: $\mathbf{U}'_p \mathbf{v} = \mathbf{U}' \mathbf{w}$, where $\mathbf{U} = [\mathbf{U}'_p, \mathbf{U}'_f]'$. Assuming that $[\mathbf{U}, \mathbf{Z}]$ is full column rank (usually, as it will be illustrated later, $[\mathbf{U}, \mathbf{Z}]$ is a selection of the columns of $\mathbf{X}$ or it is coincident with $\mathbf{X}$), and partitioning $\mathbf{D} = \text{diag}(\mathbf{D}_p, \mathbf{D}_f)$, the set of asymmetric weights minimises with respect to $\mathbf{v}$ the following objective function:

$$\varphi(\mathbf{v}) = (\mathbf{v} - \mathbf{w}_p)' \mathbf{D}_p (\mathbf{v} - \mathbf{w}_p) + \mathbf{w}'_f \mathbf{D}_f \mathbf{w}_f + [\boldsymbol{\delta}' (\mathbf{Z}'_p \mathbf{v} - \mathbf{Z}' \mathbf{w})]^2 + 2\mathbf{l}' (\mathbf{U}'_p \mathbf{v} - \mathbf{U}' \mathbf{w}).$$

The revision error arising in estimating the signal m_t is $\hat{m}_{t|t} - \hat{m}_t = \mathbf{v}' \mathbf{y}_p - \mathbf{w}' \mathbf{y}$. Replacing $\mathbf{y}_p = \mathbf{U}_p \boldsymbol{\gamma} + \mathbf{Z}_p \boldsymbol{\delta} + \boldsymbol{\epsilon}_p$, and $\mathbf{y} = \mathbf{U} \boldsymbol{\gamma} + \mathbf{Z} \boldsymbol{\delta} + \boldsymbol{\epsilon}$, and using $\mathbf{U}'_p \mathbf{v} = \mathbf{U}' \mathbf{w} = \mathbf{0}$, we obtain $\hat{m}_{t|t} - \hat{m}_t = (\mathbf{v}' \mathbf{Z}_p - \mathbf{w}' \mathbf{Z}) \boldsymbol{\delta} + \mathbf{v}' \boldsymbol{\epsilon}_p - \mathbf{w}' \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}'_p, \boldsymbol{\epsilon}'_f]'$. Hence, the first three summands of the objective function represent the mean square revision error, which is broken down into the revision error variance (the first two terms) and the squared bias term $[\boldsymbol{\delta}' (\mathbf{Z}'_p \mathbf{v} - \mathbf{Z}' \mathbf{w})]^2$. The vector $\mathbf{l}$ is a vector of Lagrange multipliers. Setting $\mathbf{Q} = \mathbf{D}_p + \mathbf{Z}_p \boldsymbol{\delta} \boldsymbol{\delta}' \mathbf{Z}'_p$, the first order conditions for the minimisation problem can be written as $\mathbf{v} = \mathbf{w}_p + \mathbf{Q}^{-1} \mathbf{Z}_p \boldsymbol{\delta} \boldsymbol{\delta}' \mathbf{Z}'_f \mathbf{w}_f - \mathbf{Q}^{-1} \mathbf{U}_p \mathbf{l}$. Premultiplying both sides for $\mathbf{U}'_p$ and recalling that $\mathbf{U}'_p (\mathbf{v} - \mathbf{w}_p) = \mathbf{U}'_f \mathbf{w}_f$, we can express the Lagrange multipliers as a linear combination of the weights $\mathbf{w}_f$. Replacing into the expression for $\mathbf{v}$ and rearranging gives

$$\mathbf{v} = \mathbf{w}_p + \mathbf{L} \mathbf{U}'_f \mathbf{w}_f + \mathbf{M} \mathbf{Z}_p \boldsymbol{\delta} \boldsymbol{\delta}' \mathbf{Z}'_f \mathbf{w}_f, \quad (1)$$

where $\mathbf{L} = \mathbf{Q}^{-1} \mathbf{U}_p [\mathbf{U}'_p \mathbf{Q}^{-1} \mathbf{U}_p]^{-1}$, $\mathbf{M} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1} \mathbf{U}_p [\mathbf{U}'_p \mathbf{Q}^{-1} \mathbf{U}_p]^{-1} \mathbf{U}'_p \mathbf{Q}^{-1}$. The matrices $\mathbf{L}$ and $\mathbf{M}$ have the following properties: $\mathbf{U}'_p \mathbf{M} = \mathbf{0}$, $\mathbf{U}'_p \mathbf{L} = \mathbf{I}$. Alternatively, the solution can be written as follows:

$$\begin{aligned} \mathbf{v} &= \mathbf{w}_p + \mathbf{L}^* \mathbf{U}'_f \mathbf{w}_f + \\ &+ \mathbf{R} \mathbf{Z}_p \boldsymbol{\delta} \boldsymbol{\delta}' [\mathbf{I} + \mathbf{Z}'_p \mathbf{R} \mathbf{Z}_p \boldsymbol{\delta} \boldsymbol{\delta}']^{-1} [\mathbf{Z}'_f - \mathbf{Z}_p \mathbf{D}_p^{-1} \mathbf{U}_p (\mathbf{U}'_p \mathbf{D}_p^{-1} \mathbf{U}_p)^{-1} \mathbf{U}'_f] \mathbf{w}_f \end{aligned} \quad (2)$$

where $\mathbf{L}^* = \mathbf{D}_p^{-1} \mathbf{U}_p (\mathbf{U}'_p \mathbf{D}_p^{-1} \mathbf{U}_p)^{-1}$, $\mathbf{R} = \mathbf{D}_p^{-1} - \mathbf{D}_p^{-1} \mathbf{U}_p (\mathbf{U}'_p \mathbf{D}_p^{-1} \mathbf{U}_p)^{-1} \mathbf{U}'_p \mathbf{D}_p^{-1}$, so that $\mathbf{U}'_p \mathbf{L}^* = \mathbf{I}$, $\mathbf{U}'_p \mathbf{R} = \mathbf{0}$. The proof of the equivalence is direct.

The DAF arise in the case $\mathbf{D} = \mathbf{K}^{-1}$ and $\mathbf{U} = \mathbf{X}$, so that the bias term is zero. On the other hand, Musgrave's asymmetric filters are obtained in the particular case when the original two-sided symmetric filter is the Henderson filter and $\mathbf{U} = \mathbf{i}$, $\mathbf{Z} = [-h, -h + 1, \dots, h]'$, $\boldsymbol{\delta} = \delta_1$, $\mathbf{D} = \sigma^2 \mathbf{I}$, that is when $\mathbf{U}$ and $\mathbf{Z}$ are respectively the first and the second column of the design matrix $\mathbf{X}$. It is nevertheless convenient for comparison purposes to reset the time origin and derive Musgrave filters applying (2) under the equivalent design $\mathbf{U} = \mathbf{i}$, $\mathbf{Z} = [1, 2, \dots, 2h + 1]'$, $\boldsymbol{\delta} = \delta_1$, $\mathbf{D} = \sigma^2 \mathbf{I}$.

3.1 The properties of the approximate filters

The approximate filters (1) raise a controversial point. The symmetric filter was derived from the assumption that the series behaves locally according to a polynomial of degree d . We seek to approximate this filter by changing our assumption about how y_t is generated, postulating that it has been possibly generated by a lower order polynomial or that the asymmetric filter is only capable of reproducing a polynomial of lower degree. In one way or another we are denying the conditions under which the original smoothing filter was derived. However, it is clear from our previous discussion that the original motivation for introducing a new class of approximating filters was the fact that the direct real time filter delivers very volatile estimates; hence, we had to move away from the direct

strategy of fitting the maintained polynomial to the available observations. Secondly, it is not implausible to assume that the behaviour of the signal at the extremes is different from that in the interior of the sampling design. An analogy can be drawn with cubic smoothing splines: the so called natural boundary condition are such that the spline is a local cubic function of time inside the boundary and is linear outside. This is beneficial to the reliability of the real time estimates and of the forecasts. The strategy that is adopted in the approximation is very similar since it effectively amounts to reducing the order of the fitting polynomial.

The merits of the class of filters (2), relative to the DAF, lie in the bias-variance trade-off. In particular, the bias can be sacrificed for improving the variance properties of the corresponding asymmetric filter. If $\mathbf{U} = \mathbf{X}$, i.e. it is asked of the filter to be capable of reproducing a d -th order polynomial, which is also the process generating the observations, the approximate filter minimising $\varphi(\mathbf{v})$ will not differ from $\mathbf{w}_a$ (in the light of the minimum distance property in section 2.2 they will coincide if $\mathbf{D} = \mathbf{K}^{-1}$); as a consequence, the real time filter will be strongly localised, and it will suffer from the same limitations as the DAF discussed in section 2.1, namely its estimates will be characterised by high variance.

When $\mathbf{U}$ is a subset of the columns of $\mathbf{X}$, spanning a polynomial of degree $d^* < d$, than we require that the filter is capable of reproducing a polynomial of degree d^* ; if the observations are generated by a polynomial of degree greater than d^* a bias will arise, which depends on the value of δ . However, the weights of the approximating filter will be more evenly distributed and the variance will be reduced. Thus, the overall mean square revision error may eventually be reduced if the actual signal is weakly evolutive.

For real time approximation, we consider the following three classes of asymmetric filters:

Asymmetric LC The asymmetric LC (Linear trend-Constant fit) real time filter arises as the best approximation to a given two-sided symmetric filter assuming that y_t is linear and imposing the constraint that the weights sum to 1, i.e. $y_{t+j} = \gamma_0 + \delta_1 j + \epsilon_{t+j}$, $\epsilon_{t+j} \sim \text{IID}(0, \sigma^2)$ and $\sum \mathbf{v}_j = 1$. So,

$$\mathbf{U} = \mathbf{i}, \mathbf{Z} = [-h, -h+1, \dots, h]', \boldsymbol{\delta} = \delta_1, \mathbf{D} = \sigma^2 \mathbf{I}.$$

The filter depends on the slope of the underlying signal through the ratio δ_1^2/σ^2 . If $\mathbf{w}$ is the symmetric Henderson filter and $\frac{\delta_1^2}{\sigma^2} = 4/(\pi R^2)$, $R = \bar{I}/\bar{C}$, then we find Musgrave filters.

Asymmetric QL The asymmetric QL (Quadratic trend-Linear fit) real time filter arises as the best approximation to a given two-sided symmetric filter assuming that y_t is quadratic and imposing the constraint that the estimates are capable of reproducing a first degree polynomial, i.e. $y_{t+j} = \gamma_0 + \gamma_1 j + \delta_2 j^2 + \epsilon_{t+j}$, $\epsilon_{t+j} \sim \text{IID}(0, \sigma^2)$ and $\sum \mathbf{v}_j = 1$, $\sum_{j=-q}^h \mathbf{v}_j j = \sum_{j=-h}^h \mathbf{w}_j j$. So,

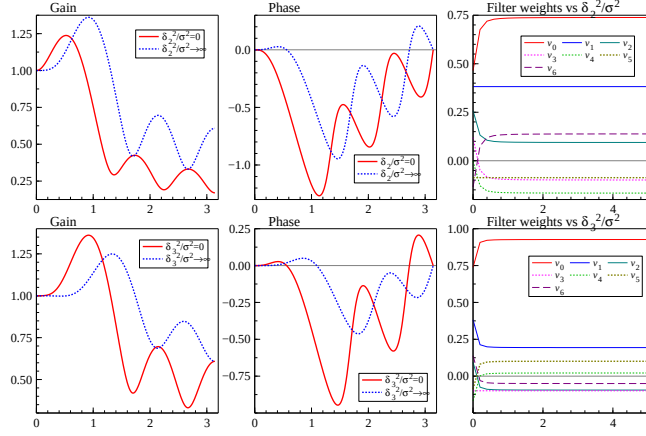
$$\mathbf{U}' = \begin{bmatrix} 1 & 1 & \dots & 1 & \dots & 1 & 1 \\ -h & -h+1 & \dots & 0 & \dots & h-1 & h \end{bmatrix}$$

$$\mathbf{Z} = [(-h)^2, (-h+1)^2, \dots, 1, 0, 1, \dots, h^2]', \boldsymbol{\delta} = \delta_2, \mathbf{D} = \sigma^2 \mathbf{I}.$$

The filter weights depend on the curvature of the signal via δ_2^2/σ^2 .

Asymmetric CQ The asymmetric CQ (Cubic trend-Quadratic fit) real time filter arises as the best approximation to a given two-sided symmetric filter assuming that y_t

Figure 2: Gain, phase and weights for the real time filter minimising the revision mean square error subject to $\sum v_i = 1$, when the observations are generated by a linear trend.



is a cubic function of time and imposing the constraint that the estimates are capable of reproducing a second degree polynomial, i.e. $y_{t+j} = \gamma_0 + \gamma_1 j + \gamma_2 j^2 + \delta_3 j^3 + \epsilon_{t+j} \epsilon_{t+j} \sim \text{IID}(0, \sigma^2)$ and $\sum v_j = 1$, $\sum_{j=-q}^h v_j j = \sum_{j=-h}^h w_j j$, $\sum v_j j^2 = \sum_{j=-h}^h w_j j^2$. So,

$$\mathbf{U}' = \begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 & 1 \\ -h & -h+1 & \cdots & 0 & \cdots & h+1 & h \\ h^2 & (-h+1)^2 & \cdots & 0 & \cdots & (h+1)^2 & h^2 \end{bmatrix},$$

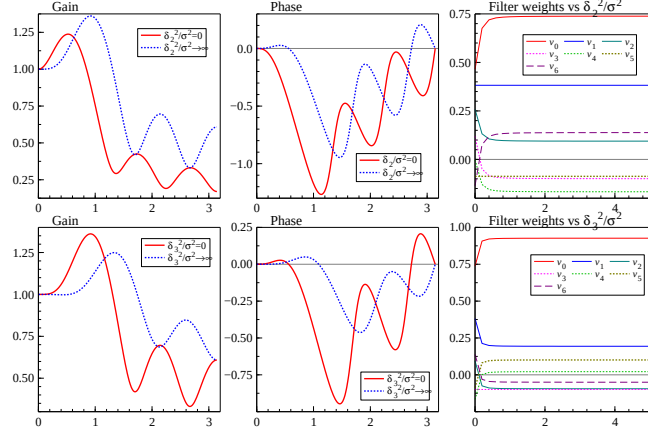
$$\mathbf{Z} = [(-h)^3, (-h+1)^3, \dots, 1, 0, 1, \dots, h^3]', \boldsymbol{\delta} = \delta_3, \mathbf{D} = \sigma^2 \mathbf{I}.$$

In this case the optimal filter depends on the parameter δ_3^2/σ^2 , which is a measure of relative inflexion.

Figure 2 plots the gain and the phase function of the asymmetric LC real time filter when $h = 6$ and $\mathbf{w}$ is the two-sided Henderson filter. As the filter depends on the slope of the underlying signal through the ratio δ_1^2/σ^2 , we plot two limiting cases arising when the slope is negligible and when it is the dominating feature. The intermediate case is the well-known Musgrave surrogate real time filter for the Henderson with 13 terms, which rises when $\delta_1^2/\sigma^2 = 4/(3.5^2\pi)$, with $R = 3.5$ being the value selected for the Henderson filter with 13 terms. The filter weights are displayed in the bottom right panel of the same figure. When the slope is negligible, $\delta_1^2/\sigma^2 = 0$ (or equivalently $R \rightarrow \infty$, which arises either when the signal is constant and is devoid of the linear term or the signal is buried in a heap of noise) the optimal approximation to the Henderson two-sided filter features weights that are less dispersed and the gain decreases from 1 almost monotonically, as it ought to be expected. The individual weights of the real time filter, $v_0, \dots, v_h$, are plotted against the value of δ_1^2/σ^2 in the bottom left panel. As the linear signal is stronger, the dispersion of the weights increases and the gain becomes higher at each individual frequencies, getting greater than one at the low frequencies.

The first row of figure 3 considers the real time QL filter and plots the gain, the phase and the individual filter weights for different values of δ_2/σ^2 , which expresses the relative importance of the curvature of the signal. The rationale for this particular type of asymmetric filters is early detection of turning points, which are a quadratic feature of

Figure 3: Gain, phase and weights for the real time filter minimising the revision mean square error: QL and CQ filters.



the signal; setting δ_2/σ^2 to an high value, the optimal filter would weight more the current observation and detect a turning point more rapidly. Essentially, with respect to the Musgrave's type of filters, the bias is reduced at the expenses of the variance.

It should also be noticed that the optimal filter for $\delta_2^2/\sigma^2 = 0$ is coincident with the optimal filter derived under a linear trend signal and using the constraint $\sum_i v_i = 0$ with $\delta_1^2/\sigma^2 \rightarrow \infty$; compare figure 2.

Finally, the bottom panels of figure 3 display the gain, phase and filter weights of the CQ real time filter. In this case the optimal filter depends on the parameter δ_3^2/σ^2 , which is a measure of relative inflexion. Again the optimal filter for $\delta_3^2/\sigma^2 = 0$ is the same as the QL filter arising for $\delta_2^2/\sigma^2 \rightarrow \infty$; compare the top panels of figure 3. As $\delta_3^2/\sigma^2 \rightarrow \infty$, the filter is the same as the direct asymmetric filter of section 2.1.

4. Illustrations

In this section we provide an illustration concerning the use of the general expression (1) for the design of real time filters suitable for a particular time series. We consider the Italian index of industrial production for the branch DL (Manufacture of electrical and optical equipment, Nace Rev. 1 classification). The reference two-sided filter is the Henderson filter and we estimate the bandwidth h by crossvalidation. The three asymmetric filters LC, QL, CQ depend on a single parameter, δ_i^2/σ^2 , $i = 1, 2, 3$. For each we compute the value that minimises the mean square revision error (MSRE), that is the value for which $\sum_{t=h+1}^{n-h} (\hat{m}_t - \hat{m}_{t|t})^2 / (n - 2h - 1)$ is a minimum.

The top left of figure 4 shows that the value of the bandwidth selected by crossvalidation is $h = 15$; the two-sided estimates of the trend are displayed in the right top panel of the same figure. We next look for the best approximation to the Henderson filter within the three particular classes. For this purpose we estimate the values of the parameters δ_i^2/σ^2 , $i = 1, 2, 3$, using a grid search. The results are presented in the bottom left panel of figure 4. The minimisers of the MSRE are $\hat{\delta}_1 = 0.103$, $\hat{\delta}_2 = 0.016$ and $\hat{\delta}_3 = 0.003$, respectively for the LC, QL and CQ classes. As illustrated by figure 4, the best approximation to the original Henderson filter is provided by the QL filter with $\hat{\delta}_2 = 0.016$. We need the real

Figure 4: Selection of a real time filter: Index of Industrial Production, Italy, branch DL. Source: Istat.

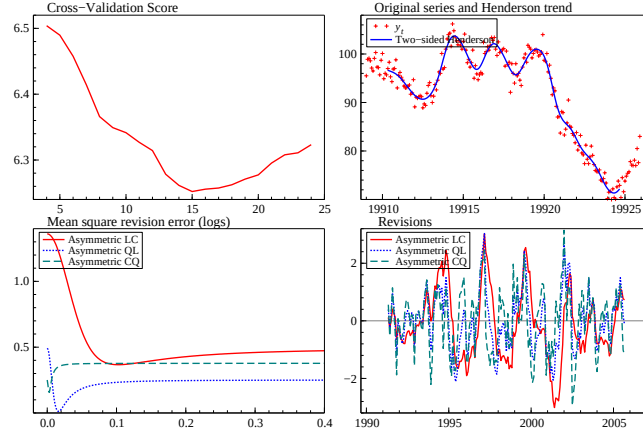
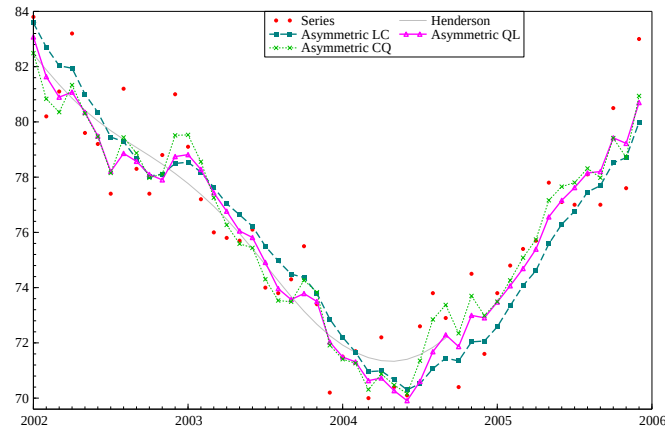


Figure 5: Index of Industrial Production, Manufacture of electrical and optical equipment, Italy. Comparison of the real time estimates arising from three approximating filters and final estimates of the trend component.



time filter to be capable of reproducing a linear signal and to react somewhat, although not in full, to the curvature of the underlying trend.

Figure 5 compares the real time estimates of the trend for the period January 2002 - December 2006, $\hat{n}_{t|t}$, arising from the best LC, QL, and CQ approximations. It is clear that the LC filter is biased when the slope is substantial: the bias is positive in a recessionary period and negative in expansion. This is so since the filter can only preserve a constant, but will distort a local linear trend. The optimal QL approximation provides the best approximation since the real time estimates are closer to the final Henderson estimates. The CQ approximation tracks the data quite well but the corresponding estimates are affected by higher variance, compared with the QL estimates.

Similar considerations apply to the DAF estimates, not reported for brevity. For the class of economic time series that are usually considered, such as industrial production, the evidence definitively points out that the direct asymmetric filter produces the most inefficient estimates, due to the very high variance inflation.

We also report the result of an extensive application to a dataset consisting of 62 industrial

production time series by market groups for the U.S. (Source: Federal Reserve Board), seasonally adjusted, available for the period January 1980 - December 2006. The final estimates of the trend are obtained by applying the Henderson filter with 13 terms ($h = 6$), which plays a prominent role in empirical applications, especially for trend estimation within the X-11 filter, appears adequate and, due to the inherent smoothness of the seasonally adjusted series, it is often coincident with the value chosen by cross-validation. The number of times the three approximations were ranked best resulted 27 for the LC, 34 for the QL and only 1 for the CQ filter.

The evidence presented in this section illustrates that the proposal of designing asymmetric filters in a more general and flexible way helps estimating the underlying signal with more accuracy.

5. Conclusions

The paper has considered the problem of estimating the trend of a time series in real time by means of local polynomial filters; we showed that automatic adaptation at the boundary fails due to the high volatility of the estimates. We thus evaluated the strategy of approximating a given symmetric local polynomial filter by minimising the mean square revision error subject to different order polynomial reproducing constraints and by making certain assumptions concerning the nature of the underlying signals. Restricting our attention to three families of real time filters that depend on certain key feature of the unknown signal, such as its slope and curvature, we proposed to estimate these key feature from the available data, rather than taking a fixed filter.

Our empirical illustrations concerned the minimum mean square revision error approximation of the Henderson filter, a very popular local cubic smoother. They enable us to conclude that we can improve upon the well-known Musgrave asymmetric filters, which for the series considered suffers from very large revisions especially in steep recessions and recoveries and around turning points. This evidence arises as a consequence of the fact that the filter is not designed to deal with signals characterised by strong slope and curvature.

We also considered the strategy of building either direct or minimum revision mean square approximations using a nearest neighbour bandwidth, but its effectiveness in ameliorating the approximation to the Henderson filter was not proven by our empirical applications.

References

- Berlinet A. and Thomas-Agnan C. (2004) *Reproducing Kernel Hilbert Space in Probability and Statistics*, Kluwer Academic Publisher.
- Doherty M. (2001) Surrogate henderson filters in x-11., *Australian and New Zealand Journal of Statistics*, 43, 385–392.
- Gray A. and Thomson P. (2002) On a family of finite moving-average trend filters for the ends of series, *Journal of Forecasting*, 21, 125–149.
- Henderson R. (1916) Note on graduation by adjusted average, *Transaction of the Actuarial Society of America*, 17, 43–48.
- Musgrave J. (1964) *A Set of End Weights to End All End Weights*, Working paper. Census Bureau, Washington.