Kolmogorov-Wiener Filters for Finite Time-Series

Christoph Schleicher*†

First version: March 2002
This version: October 2002
Compiled: January 15, 2003

Abstract

This paper describes a framework of how to optimally implement linear filters for finite time series. The filters under consideration have the property that they minimize the mean squared error compared to some ideal hypothetical filter. It is shown in examples that three commonly used filters, the bandpass filter, the Hodrick-Prescott filter and the digital Butterworth filter need to be adjusted when applied to finite samples of serially correlated or integrated data. An empirical example indicates that the proposed optimal filters improve the end-of-sample performance of standard filters when applied to U.S. GDP data.

Keywords: business cycles, mechanical filters, spectral analysis, bootstrap
JEL classification: C22

* Department of Economics, University of British Columbia, #997 - 1873 East Mall Vancouver, BC Canada V6T-1Z1. Phone: (604) 822-5661, fax: (604) 822-5915, email: christoph@iam.ubc.ca.

† I would like to thank Paul Beaudry, Henryk Gurgul, Theodore Kolokolnikov, James Nason, Anthony Peirce, and Simon van Norden for helpful comments. All remaining mistakes are my responsibility. An earlier version of this paper was presented at the Canadian Economics Association Conference (Calgary, June 2002).

1 Introduction

Linear filters are ubiquitous in applied macroeconomic research, ranging from simple differencing operations, mechanical detrending devices, seasonal adjustment to ARIMA models. Many of these filters have very long and often infinite impulse response sequences, such that some approximation procedure is necessary for finite samples. This paper shows that the time-series properties of the data play an important role in such finite sample approximations.

Two different schools of thought stand out in the existing literature on filtering of economic time-series. Members of the first group define a statistical model, usually consisting of a trend component, a cyclical component and noise, which the underlying economic time-series is assumed to follow. The optimal Wiener filter is then the signal extracting device, which minimizes the mean squared error to the actual data generating process. Examples for filters based on such statistical models are the exponential smoothing filter used by Lucas (1980), the Hodrick-Prescott filter (1997) and more recently the Butterworth filter known from electrical engineering (Pollock, 1999 and 2000). It is important to note that these filters are based on a hypothetical data generating process of infinite length, such that the properties of the filters have to interpreted in an asymptotic context, as opposed to their finite sample counterparts.

Members of the second school assume the existence of an ideal filter, often characterized in terms of its frequency domain representation (transfer function), which is then used as a benchmark for finite sample approximations. The Kolmogorov-Wiener filter is defined as the finite sample filter, which minimizes the mean-squared error to the ideal filter. Often the ideal filter is assumed to be an ideal bandpass filter with a square-wave transfer function, an assumption that is implicit in the work of Baxter and King (1999), Christiano and Fitzgerald (1999) and Pedersen (2001) and, to a certain extent, in Guay and St-Amant (1995) and Pollock (2000). Pedersen, for example, constructs a metric which is essentially a numerical approximation to the L^2 distance of a filter to the ideal bandpass filter. The choice of the ideal bandpass filter as a reference model is justified on grounds of the Burns and Mitchell (1946) taxonomy, which classifies business cycles as components with cycles between 8 and 32 quarters.

The approach in this paper lies in the neutral ground shared by both schools of thought mentioned above. It is agnostic, in the sense that no statements are made about why a certain filter (either based on an underlying statistical model or the shape of its transfer function) should be preferred over another. Instead it is assumed that the ideal filter is known and the focus is shifted towards an optimal approximation for a finite sample of data that is possibly serially correlated and nonstationary.

The methodology used is similar in spirit to the work by Christiano and Fitzgerald (1999) on approximations to the ideal bandpass filter. It is shown that an extended framework applies to a wide class of commonly used filters with a linear impulse response representation. Besides the ideal bandpass filter, examples discussed in this paper include the Hodrick-Prescott filter and the Butterworth digital filter.

The analysis in this paper is based on Hilbert space methodology familiar from time-series econometrics. Optimal use of limited information corresponds to projecting the possibly infinite convolution product of the ideal filter onto a finite subspace, which is spanned by the data sample. The fundamentals of the theory of prediction and interpolation of time-series were developed independently by Kolmogorov (1941) and Wiener (1949). While Kolmogorov's analysis was based on the time domain, Wiener used frequency domain methods. Accordingly, the optimization problem for the finite sample approximation can be stated both in the time- and the frequency domain. The latter approach is appealing from an intuitive point of view, since the spectral density of the underlying time-series effectively serves as a weighing function for the accuracy of the finite sample approximation. For negatively autocorrelated processes with strong mean reversion, the focus is shifted to the high frequency components, while for integrated time-series accuracy is most important at the lowest frequencies. The results are summarized for stationary time-series in proposition 1 and for nonstationary time-series in propositions 2 to 4. A consequence is that the same filter can produce varying results when applied to different time series. Cogley and Nason (1995) discuss this phenomenon in the case of the Hodrick-Prescott filter. A custom-tailored approach of filter-design is therefore desirable to achieve comparable and consistent results.

Section 2 of the paper discusses the underlying theoretical framework of optimal filters for finite time series and proposes solutions for a range of commonly used time-series models as well as directions of how to implement them. Section 3 applies the methodology to three commonly used filters, the bandpass filter, the Hodrick-Prescott filter and the Butterworth filter. A comparison is made withe previous solution approaches. Section 4 compares how different implementations affect the performance of filters at the end of the sample. Section 5 concludes.

2 Finite filters and time series

This section introduces the basic problem of optimal finite sample approximation and subsequently proposes solutions for stationary and integrated time series.

2.1 The filtering problem in the time and frequency domain

Let us assume that we are interested in the linear transformation

$$\begin{aligned} y_t &= B(L)x_t \\ &= \sum_{j=-\infty}^{\infty} B_j x_{t+j}, \end{aligned} \tag{1}$$

where $\{B_j\}_{j=-\infty}^{\infty}$ is the impulse response sequence of some ideal linear filter. In finite samples this transformation is not feasible, in general, and it is necessary to use an alternative filter with finite impulse sequence $\{\hat{B}_{t,j}\}_{j=-n_1}^{n_2}$,

$$\begin{aligned} \hat{y}_t &= \hat{B}_t(L)x_t \\ &= \sum_{j=-n_1}^{n_2} \hat{B}_{t,j} x_{t+j}. \end{aligned} \tag{2}$$

This finite sample filter does not have to be neither symmetric (i.e. $\hat{B}_{t,j} = \hat{B}_{t,-j}$) nor time-invariant (i.e. $\hat{B}_{t,j} = \hat{B}_j$). The only restrictions we need to impose is that for a sample with T observations, $n_1 < t$ and $n_1 + n_2 + 1 \equiv N \leq T$. Let $J = -n_1, ..., n_2$ denote the index set (or information set) of $\hat{B}_t$, then the restriction claims that $J \oplus t \subset \{1, ..., T\}$. In other words, y_t is a linear combination of a subset of $\{x_t\}_{t=1}^{T}$ such that $y_t \in \overline{\text{sp}}(\{x_j\}_{J\oplus t})$. In the line of Kolmogorov (1941) and Wiener (1949) we are looking for a sequence $\{\hat{B}_{t,j}\}$ that minimizes the mean squared error between y_t and $\hat{y}_t$,

$$\{\hat{B}_{t,j}\} = \text{argmin } \mathrm{E}[y_t - \hat{y}_t]^2. \tag{3}$$

To restate the problem more accurately, assume that $\{x_t\}$ is the realization of a stochastic process $\{X_t\}$, defined on the real probability space $L^2(\Omega, \boldsymbol{F}, P)$ with finite inner product

$$\mathrm{E}(XY) = \langle X, Y \rangle = \int_\Omega X(\bar{\omega})Y(\bar{\omega})dP(\bar{\omega}) < \infty. \tag{4}$$

Then

$$\{\hat{B}_{t,j}\} = \text{argmin } \int_\Omega \left[y_t(B, X(\bar{\omega})) - \hat{y}_t(\hat{B}_t, X(\bar{\omega}))\right]^2 dP(\bar{\omega}). \tag{5}$$

This definition is inconvenient to work with and matters become more tractable when the problem is transformed into the frequency domain $L^2([-\pi, \pi], \boldsymbol{B}, \mu)$ using the isomorphic mapping $IX_t = e^{it}$ (the Fourier transform)[1]. The Fourier transform has the opportune property that convolution in the time domain becomes simple multiplication in the frequency domain, such that we can reexpress equations (1) and (2) as

$$y(\omega) = B(\omega)x(\omega), \quad \omega \in [-\pi, \pi] \tag{6}$$

and

$$\hat{y}(\omega) = \hat{B}(\omega)x(\omega), \quad \omega \in [-\pi, \pi], \tag{7}$$

[1]The frequency domain is defined on the interval $[-\pi, \pi]$ and its corresponding Borel σ-field. μ is the uniform probability measure $d\mu(\omega) = \frac{1}{2\pi}d\omega$ that normalizes the inner product

$$\langle IX, IY \rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)\bar{Y}(\omega)d\omega.$$

The Fourier transform is the pair of mappings

$$f(\omega) = \sum_{-\infty}^{\infty} f_j e^{i\omega j} \qquad \text{(analysis equation)}$$

$$f_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)e^{-i\omega j}d\omega \qquad \text{(synthesis equation)} \quad .$$

where $B(\omega)$ and $\hat{B}(\omega)$ are the transfer functions of the filters B and $\hat{B}$. The transfer function provides an useful tool to model a filter's performance in the frequency domain, and is usually written in polar form as

$$H(\omega) = \Gamma(\omega)e^{i\theta(\omega)}.$$

$\Gamma(\omega) = |(H(\omega)|$ is the filter's gain function and determines how the amplitude of a time series is increased or diminished at frequency ω. $\theta(\omega) = \arg(H(\omega))$ is the filter's phase function and determines how cycles at frequency ω are shifted forward or backward in time. Taking squared norms in (6) we see that the squared gain function gives a direct relation between the spectral densities $f_y(\omega)$ and $f_x(\omega)$:

$$f_y(\omega) = \|y(\omega)\|^2 = |B(\omega)|^2\|x(\omega)\|^2 = |B(\omega)|^2 f_x(\omega) \tag{8}$$

Using the properties of isomorphisms (see for example Brockwell and Davis (1991) we can write

$$\begin{aligned}
\mathrm{E}[y_t - \hat{y}_t]^2 &= \|y_t - \hat{y}_t\|^2 \\
&= \|I(y_t - \hat{y}_t)\|^2 = \|Iy_t - I\hat{y}_t\|^2 = \|y(\omega) - \hat{y}(\omega)\|^2 \\
&= \|[B(\omega) - \hat{B}(\omega)]x(\omega)\|^2 \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|B(\omega) - \hat{B}(\omega)\right|^2 f_x(\omega)d\omega. \qquad (9)
\end{aligned}$$

Equation (9) provides an alternative and intuitive interpretation to the mean squared error minimization problem. In the frequency domain the optimal filter minimizes (the square of) the L^2 norm of the difference between the transfer functions, weighed by the spectral density of the underlying time series.

2.2 Stationary interdependence

In this section we are concerned with stochastic processes $\{X_t\}$ that are stationary in the sense that

$$\begin{aligned}
&\mathrm{E}X_t = \mu, \\
&\mathrm{E}(X_t - \mu)^2 = \gamma_0 < \infty, \\
&\mathrm{E}(X_t - \mu)(X_{t-k} - \mu) = \mathrm{E}(X_s - \mu)(X_{s-k} - \mu) = \gamma_k \quad \text{for all } t, s \in \mathbb{Z}.
\end{aligned}$$

We will further assume that $\lim_{k\to\infty} \gamma_k = 0$, such that $\{X_t\}$ is the class of all stationary ARMA models

$$\phi(L)X_t = c_1 + \theta(L)\epsilon_t, \quad \epsilon \sim \mathrm{WN}, \tag{10}$$

with Wold representation

$$X_t = c_2 + \frac{\theta(L)}{\phi(L)}\epsilon_t = c_2 + \psi(L)\epsilon_t. \tag{11}$$

The optimization problem

$$\{\hat{B}_j\} = \mathrm{argmin}\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|B(\omega) - \hat{B}(\omega)\right|^2 f_x(\omega)d\omega \tag{12}$$

can be viewed as a projection of $B(\omega)f_x(\omega)$ onto the subspace $\boldsymbol{M}_{\hat{B}} = \overline{\mathrm{sp}}\{e^{i\omega j}, j \in J\}$ of $L^2([-\pi, \pi])$. The frequency domain itself is spanned be the set of all Fourier basis functions $\{e^{i\omega j}, j \in \mathbb{Z}\}$, which form an orthonormal basis. In the absence of serial correlation the spectrum is flat and the optimal finite sample filter is just a Dirichlet window of the ideal filter ($\hat{B}_j = B_j, j \in J$). In this case simple truncation provides the best approximation of the ideal filter in a finite sample. If, on the other hand, $\gamma_k \neq 0$ for some $k \leq Q \in \mathbb{N}$ and $\gamma_k = 0$ for $k > Q$, consider the spectrum defined as the Fourier transform (Wiener, 1930 and Khintchine, 1934)

$$f_x(\omega) = \sum_{k=-Q}^{Q} \gamma_k e^{i\omega k}. \tag{13}$$

If the coefficients of B are square summable, the integral in (12) is guaranteed to be finite, since

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|B(\omega) - \hat{B}(\omega)\right|^2 f_x(\omega)d\omega \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |B(\omega)|^2 f_x(\omega)d\omega \le \sup_{\omega} f_x(\omega) \sum_{-\infty}^{\infty} B_j^2,$$

(the spectrum is finite because of stationarity). When $B(\omega)$ is multiplied by the spectrum, its Fourier coefficients are effectively remixed to account for serial correlation, such that the optimal final sample approximation becomes a linear combination of the ideal filter sequence, weighed by the autocovariance function. The exact definition of the optimal finite sample filter is given in the following proposition.

Proposition 1 *(Stationary Interdependence)*

Let $\{X_t\}$ be a stationary process with autocovariance function γ_k and $\gamma_k = 0$ for $k > Q$. Then its optimal finite sample filter sequence $\hat{B} = [\hat{B}_{-n_1}, ..., \hat{B}_{n_2}]'$ (with respect to a square-summable ideal filter) is given by[2]

$$\hat{B} = \hat{\mathbf{\Gamma}}^{-1}\mathbf{\Gamma}B, \tag{14}$$

where $\hat{\mathbf{\Gamma}}$ is an $(N) \times (N)$ (Toeplitz) matrix with typical element

$$\hat{\mathbf{\Gamma}}_{m,n} = \gamma_{|m-n|}, \tag{15}$$

[2]In practice direct matrix inversion can be avoided in the solution of (14), since $\hat{\mathbf{\Gamma}}$ is a band-diagonal Toeplitz matrix, which can be factored by Cholesky decomposition (see Pollock 1999, p. 608).

$\mathbf{\Gamma}$ *is an* $(N) \times (N + 2Q)$*-matrix with typical element*[3]

$$\mathbf{\Gamma}_{m,n} = \gamma_{|m+Q-n|}, \tag{16}$$

and $B = [B_{-n_1-Q}, ..., B_{n_2+Q}]'$ *is a column vector of coefficients of the ideal filter.*

Proof: Appendix A.

2.3 Integrated processes

It is widely accepted that the assumption of stationarity applies to the growth rate rather than the level of most macroeconomic and many financial variables. Often these series are characterized by Granger's (1966) typical spectrum shape, with most of the power concentrated in the low frequencies. These observations lead to the family of ARIMA models, of which the pure random walk

$$(1 - L)X_t = \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2), \tag{17}$$

with spectral density

$$f_x(\omega) = \frac{\sigma^2}{|1 - e^{i\omega}|^2}, \tag{18}$$

is a canonical example. Compared to the stationary case, the analysis is complicated by the fact that the random walk has infinite variance, such that an additional restriction has to be imposed on the optimal filter in order to remain in the familiar L^2 space. As Christiano and Fitzgerald (1999) show for the ideal bandpass filter, this restriction removes the double pole at zero frequency in the spectrum of the random walk. This is equivalent to first integrating the approximate and ideal filter sequences,

[3] For the case when $Q = \infty$, $\mathbf{\Gamma} B_j$ represents the infinite sums $\sum_{k=-\infty}^{\infty} \gamma_{|k-j|} B_k$, $(j = -n_1, ..., n_2)$.

taking first order conditions and then taking first differences again. The result has a strikingly simple form: All optimal finite sample filter coefficients are identical to their ideal filter counterparts, except the first and last element, which are equal to the sum of the (truncated) lower and upper tail of the ideal filter sequence, respectively:

$$\hat{B}_j = B_j \quad \text{for} \quad j = -n_1 + 1, ..., n_2 - 1,$$
$$\hat{B}_{n_1} = \textstyle\sum_{k=-\infty}^{-n_1} B_k, \quad \hat{B}_{n_2} = \textstyle\sum_{k=n_2}^{\infty} B_k. \tag{19}$$

Proposition 2 *(Integration)*

If $\{X_t\}$ *follows a random walk*

$$\Delta X_t = \epsilon_t, \qquad \epsilon_t \sim \text{WN}(0, \sigma^2),$$

and the optimal filter $\{B_j\}_{j=-\infty}^{\infty}$ *satisfies the conditions*

(i) $\sum_{j=-\infty}^{0} |\sum_{k=-\infty}^{j} B_k|^2 < \infty$, *and*

(ii) $\sum_{j=1}^{\infty} |\sum_{k=j+1}^{\infty} B_k|^2 < \infty$,

the optimal finite sample filter approximation $\hat{B} = [\hat{B}_{-n_1}, ..., \hat{B}_{n_2}]'$ *is given by*

$$\hat{B} = \begin{bmatrix} \boldsymbol{D} \\ \iota \end{bmatrix}^{-1} \begin{bmatrix} C \\ \beta \end{bmatrix} \tag{20}$$

where $\boldsymbol{D}$ *is an* $(N-1) \times N$ *matrix whose first* $N-1$ *columns and rows constitute a lower triangular matrix filled with ones and zero entries in the last column,* ι *is an* $1 \times N$*-vector of ones,* C *is the* $N-1$ *column vector* $C = [C_{-n_1}, ..., C_{n_2-1}]'$*, whose elements are given by* $C_j = \sum_{k=-\infty}^{j} B_k$ *and* $\beta = \sum_{j=-\infty}^{\infty} B_j$*.*

Proof: Appendix B.

In general closed form solutions for the infinite sums in proposition 2 are not available, with the important exception of symmetric (even) and anti-symmetric (odd) filters. Fortunately, the filters that are most widely used in economic research and in particular the filters discussed subsequently in section 3 (the Hodrick-Prescott filter, approximate bandpass filters and the Butterworth filter) are symmetric and have optimal finite sample approximations with closed form matrix representations, as described in the following proposition.

Proposition 3 *(Integration and Symmetric Filters)*

If $\{X_t\}$ follows a random walk and the ideal filter B has an even impulse response sequence ($B_j = B_{-j}$), the optimal finite sample filter approximation can be computed as

$$\hat{B} = \begin{bmatrix} \boldsymbol{D} \\ \iota \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{M}B + \frac{\beta}{2}\tau \\ \beta \end{bmatrix} \tag{21}$$

where $\boldsymbol{M}$ is the $(N-1) \times N$ matrix defined as

$$\boldsymbol{M} = \begin{bmatrix} \underset{n_1 \times 1}{\mathbf{0}} & \underset{n_1 \times n_1}{\boldsymbol{M}_1} & \underset{n_1 \times n_2}{\mathbf{0}} \\ \underset{n_2 \times n_1}{\mathbf{0}} & \underset{n_2 \times n_2}{\boldsymbol{M}_2} & \underset{n_2 \times 1}{\mathbf{0}} \end{bmatrix}$$

where $\boldsymbol{M}_1$ is an upper triangular matrix, filled with $-\frac{1}{2}$ in the last column and -1 in all other columns. $\boldsymbol{M}_2$ is a lower triangular matrix, filled with $\frac{1}{2}$ in the first column and 1 in all other columns. τ is an $(N-1) \times 1$ vector of ones and $\hat{B}$ and B are N-vectors of the filter coefficients.

Proof: Appendix C.

For symmetric filters with high-pass properties ($\beta = 0$) the filter coefficients are given by the following simple expression:

$$\hat{B}_j = B_j \quad \text{for } j = -n_1+1, ..., n_2-1,$$
$$\hat{B}_{-n_1} = \tfrac{B_0}{2} - \textstyle\sum_0^{n_1-1} B_j, \qquad \hat{B}_{n_2} = \tfrac{B_0}{2} - \textstyle\sum_0^{n_2-1} B_j. \tag{22}$$

By combining the results of propositions 1 and 2 the analysis can be extended to the class of ARIMA models

$$\phi(L)\Delta X_t = c_1 + \theta(L)\epsilon_t, \quad \epsilon \sim \text{WN}, \tag{23}$$

whose stationary component has a Wold representation

$$\Delta X_t = c_2 + \frac{\theta(L)}{\phi(L)}\epsilon_t = c_2 + \psi(L)\epsilon_t. \tag{24}$$

Proposition 4 *(ARIMA models)*

If $\{X_t\}$ is an integrated process with a stationary component with autocovariance function γ_k ($\gamma_k = 0$ for $k > Q$), its optimal finite sample approximation is given by

$$\hat{B} = \begin{bmatrix} \hat{\mathbf{\Gamma}}\boldsymbol{D} \\ \iota \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Gamma}C \\ \beta \end{bmatrix} \tag{25}$$

where $\hat{\mathbf{\Gamma}}$ and $\mathbf{\Gamma}$ are $(N-1)\times(N-1)$ and $(N-1)\times(N-1+2Q)$ variations of the matrices introduced in proposition 1 and C is the $N-1+2Q$-vector $[C_{-n_1-Q}, ..., C_{n_2-1+Q}]'$ variation of the integrated coefficients vector from proposition 2.

In the special case where B is symmetric, the optimal finite sample filter can be computed as

$$\hat{B} = \begin{bmatrix} \hat{\mathbf{\Gamma}}\boldsymbol{D} \\ \iota \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Gamma}(\boldsymbol{M}B + \frac{\beta}{2}\tau) \\ \beta \end{bmatrix} \tag{26}$$

where $\boldsymbol{D}$ *and* $\boldsymbol{M}$ *are* $(N-1)\times N$ *and* $(N-1+2Q)\times(N+2Q)$ *variations of the matrices from proposition 3.* $\boldsymbol{M}$ *is modified as*

$$\boldsymbol{M}=\left[\begin{array}{ccc} \underset{(n_1+Q)\times 1}{\mathbf{0}} & \underset{(n_1+Q)\times(n_1+Q)}{\boldsymbol{M}_1} & \underset{(n_1+Q)\times(n_2+Q)}{\mathbf{0}} \\ \underset{(n_2+Q)\times(n_1+Q)}{\mathbf{0}} & \underset{(n_2+Q)\times(n_2+Q)}{\boldsymbol{M}_2} & \underset{(n_2+Q)\times 1}{\mathbf{0}} \end{array}\right].$$

ι *is an* N*-vector filled with ones,* τ *is an* $(N-1+2Q)$*-vector whose first* $Q+N-1$ *entries equal* 1 *and the last* Q *entries equal* -1*, and* B *is the* $N+2Q$*-vector* $[B_{-n_1-Q},...,B_{n_2+Q}]'$.

Proof: Combining the proofs of propositions 2 and 3.

2.4 Implementation of the optimal filter

In practice the construction of the optimal filter consists of two separate problems, the estimation of the time-series properties of the filtered signal and the construction of the corresponding optimal finite impulse response sequences discussed in the previous sections.

The first step involves determining whether the time-series is stationary or contains an integrated trend component. If the series is stationary, the autocovariance function can be computed either by directly estimating the second sample moments, or by fitting an ARMA model and computing the autocovariance function analytically. If the time series includes an autoregressive component, the autocovariance function dies out at a geometric rate and some cutoff value has to be taken. This truncation is unlikely to seriously affect the accuracy of the filter unless the time-series is nearly integrated, in which case the assumption of a random walk may be a better approximation. If the time series follows an ARIMA model, the optimal solution is based on the autocovariance function of the first differences of the series, which, again, can be estimated directly from the data, or by first fitting a parametric model.

For the second problem it is necessary to compute the coefficients of the filter's impulse response function. This is not always straightforward since many filters are described by their frequency response function $B(e^{-i\omega})$ rather than by their time domain representation. Examples include the bandpass filter, the Hodrick-Prescott filter and the Butterworth filter, which are discussed in the following section. In this case the filter coefficients can be computed via the inverse Fourier transform

$$B_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} B(e^{-i\omega}) e^{i\omega j} d\omega. \tag{27}$$

In the case of the bandpass filter the analytical solution is straightforward, however in the case of the Hodrick-Prescott filter and the Butterworth filter it involves infinite sums, such that numerical integration is the preferred method of computation.

Another issue arises if the ideal filter has an infinite impulse response function that is neither symmetric nor anti-symmetric and the underlying signal is integrated. In this case the infinite sums of the optimal solution as in equation (19) have to be approximated by finite sums. In terms of the overall performance the accuracy of these partial sums is likely to be less important than the rule that all coefficients sum up to zero. A heuristic approach is then to first compute coefficients $\tilde{B}_j$ and then subtract the N-th fraction of their sum

$$\hat{B}_j = \tilde{B}_j - \frac{1}{N} \sum_{k=-n1}^{n_2} \tilde{B}_k. \tag{28}$$

This method is used, for example, by Baxter and King (1999) for approximate band-pass filters for integrated time-series.

3 Three commonly used filters

This section gives examples of optimal finite sample approximations for three different filter models. The first two, the ideal bandpass filter and the Hodrick-Prescott filter have been used widely in applied economic research. The third, the digital Butterworth filter, is well known in electrical engineering and audio-acoustic research and has been recently proposed by Pollock (1999 and 2000) as a detrending device for economic time-series.

3.1 Approximate bandpass filters

Bandpass filters are a standard tool in digital signal processing. For an ideal bandpass filter the transfer function equals a square wave, it is set equal to one for frequencies within the passband (the interval $(a, b]$) and equal to zero for all other frequencies (the stopband),

$$B(e^{-i\omega}) = \begin{cases} 1, & \text{if } |\omega| \in (a, b] \\ 0, & \text{otherwise.} \end{cases} \tag{29}$$

The impulse response coefficients of the ideal lowpass filter can be found by applying the inverse Fourier transform to the transfer function $B(e^{-i\omega})$

$$\begin{aligned} B_0 &= \frac{b-a}{\pi} \\ B_j &= \frac{\sin(bj) - \sin(aj)}{\pi j}. \end{aligned} \tag{30}$$

Two observations can be made at this point: First, since the sine function is an even function, the impulse response function of an ideal bandpass filter is symmetric about the origin, therefore the phase effect is zero over the entire frequency band. Second, the sequence of filter coefficients extends indefinitely in both directions. Therefore, in praxis only approximations to the ideal bandpass filter are possible. The simplest approach is to truncate the infinite ideal filter sequence with the Dirichlet kernel. The resulting finite sample filter has a transfer function that is a finite Fourier series approximation to the ideal square wave function. A prominent feature of such a finite sample transfer function is a series of oscillations around the transition from stopband to passband, which is referred to as the Gibbs (1899) effect in the engineering literature.

It is easy to verify, that the Dirichlet window provides the optimal finite sample approximation in the case of serially uncorrelated time-series. For example, Baxter and King (1999) recommend a time-invariant symmetric Dirichlet window of 25 coefficients for quarterly, and a Dirichlet window of 7 coefficients for yearly stationary macroeconomic data. For integrated data Baxter and King construct an alternative filter by subtracting the mean from the Dirichlet window such that $\hat{B}(1) = 0$ (highpass condition). Christiano and Fitzgerald (1998) show that the optimal approximation for integrated time-series is given by proposition 3 rather than the filter proposed by Baxter and King

$$\begin{gathered}\hat{B}_j = B_j \quad \text{for } j = -n_1+1, ..., n_2-1, \\ \hat{B}_{-n_1} = \tfrac{B_0}{2} - \textstyle\sum_0^{n_1-1} B_j, \qquad \hat{B}_{n_2} = \tfrac{B_0}{2} - \textstyle\sum_0^{n_2-1} B_j. \end{gathered} \tag{31}$$

They further advocate the use of larger windows and question the necessity to restrict oneself to symmetric zero phase filters. In fact they show that the Kolmogorov-Wiener approximation always makes use of the full sample, even at the beginning and the end of the time-series, and suggest that the phase effect of asymmetric filters is less serious than previously assumed.

The following example illustrates how an optimal finite sample approximation can be constructed for a process with strong mean reversion.

Example 1: ARMA(1,1): Consider the following ARMA(1,1) process

$$(1 + 0.9L)x_t = (1 - 0.3L)\epsilon_t \qquad \epsilon_t \sim \text{WN}(0, 1).$$

As is shown in the first graph of figure (1), almost all of the spectral mass is concentrated at the highest frequencies. This is an example for which the Dirichlet approximation of the stationary Baxter-King filter as well as the I(1) filter recommended by Christiano and Fitzgerald does not work well, since the autocorrelation is strongly negative. Accordingly the optimal filter (second row, left) minimizes leakage close to $\pm\pi$, but shows relatively little concern about discrepancies at the low frequencies. In this example the filter-length is chosen to as 3 on each side ($n_1 = n_2 = 3$) and the pass-band is between $\pi/6$ and $\pi/2$. Numerical integration over the squared leakage, weighted by spectral density ($\equiv$ effective leakage), indicates that the penalty function (squared difference of L^2-norm) is almost two times lower for the optimal filter than for the standard *iid*-filter. To obtain the optimal filter, first calculate the Wold representation ($\psi_0 = 1, \psi_1 = \theta + \phi = -1.2, \psi_k = \psi_1\phi^{(k-1)}, \quad k = 1, 2, ...,$ where $\phi = -0.9$ and $\theta = -0.3$). The autocovariance function is then given by

$$\gamma_k = \begin{cases} 1 + \sum_{j=0}^{\infty} \psi_1^2 \phi^{2j} & k = 1, \\ \psi_1 \phi^{k-1} + \sum_{j=0}^{\infty} \psi_1^2 \phi^{2j+k} & k > 1. \end{cases}$$

Using γ_k it is straightforward to construct the matrices $\mathbf{\Gamma}$ and $\hat{\mathbf{\Gamma}}$ from proposition 1. The autoregressive component of the model has infinite memory, but decays at a geometric rate. Setting the cutoff parameter Q to 100 accounts for all components that are larger than 10^{-4}.

Example 2: I(1) process: If the underlying process is integrated of order 1, its spectrum approaches infinity at the origin, hence an optimal filter concentrates all accuracy towards the lowest frequencies. Figure (2) compares the performance of an iid filter (Dirichlet approximation) with the optimal filter from proposition 2 (Christiano-Fitzgerald filter). Because the transfer function of the iid filter does not equal zero at the origin, the criterion of the minimization problem (12) becomes infinity. An important observation of figure (2) is that even the relatively small leakage of the optimal filter becomes large in the proximity of the origin when it is multiplied by the spectrum. As a result, the filter generates cycles that lie outside the desired bandpass, an effect that is amply discussed in Cogley and Nason (1995a), Guay and St-Amant (1996) and Pedersen (2001). The only remedy against this problem is to use a filter window, which is as large as possible, which is what Christiano and Fitzgerald propose.

Figure 3 compares the Dirichlet truncation and the optimal filter when applied to a sample of 29 observations, with a random walk as the underlying data generating process, a frequent scenario in applied macroeconomic research. In the middle of the sample the transfer functions of both filters are similar and exhibit the Gibbs effect, which is typical for truncated bandpass filters. However, the small amount of leakage of the Dirichlet filter near the zero frequency is greatly amplified by the singularity of the spectral density of the random walk. Note also, that the phase effect is zero[4]. At the end of the sample the transfer functions of the two filters differ considerably, at first sight both the gain function and especially the transfer function of the optimal filter are inferior to that of the Dirichlet filter. The superior performance becomes

[4]The jumps of size π in the phase spectrum are due to the fact that the transfer function is negative, but the gain function is defined as a positive real function.

clear only when the squared deviation of the transfer function is weighted by the spectral density of the random walk.

3.2 The Hodrick-Prescott filter

The Hodrick-Prescott filter is a linear time-invariant symmetric filter with the infinite moving average representation

$$HP(L) = \frac{\lambda(1-L)^2(1-L^{-1})^2}{\lambda(1-L)^2(1-L^{-1})^2+1}, \tag{32}$$

It also acts as a high-pass filter, since the transfer function equals zero at zero frequency and approaches unity at π radians. The parameter λ simultaneously determines the cutoff frequency and the rate of transition of the gain function. Since it has 4 zeros at zero frequency HP filter renders series that are integrated up to order 4 stationary (King and Rebelo 1993). However, since the denominator also includes the terms $\lambda(1-L)^2(1-L^{-1})^2$, these differences are partly undone, so that more low frequency components are still passed through than for example for the first difference filter. The Hodrick-Prescott filter can be viewed as a special case of the family of smoothing splines developed by Reinsch (1976), in the sense that it is the asymptotic solution (as $T \to \infty$) to the following optimization problem, which minimizes the squared deviations of a time series from its trend subject to a smoothness constraint

$$\min_{\{s_t\}_{s=1}^{T}} \sum_{t=1}^{T}(y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1}[(s_{t+1}-s_t)-(s_t-s_{t-1})]^2. \tag{33}$$

The first order conditions for this problem comprise a system of T variables in T unknowns, and the cyclical component c_t can be recovered as

$$c = (\mathbf{I}_T - \mathbf{M}^{-1})y, \tag{34}$$

where c and y are T-vectors, $\mathbf{I}_T$ is a $T-$identity matrix and $\mathbf{M}$ is a Toeplitz matrix with diagonal band $[\lambda, -4\lambda, 1+6\lambda, -4\lambda, \lambda]$, initial and end conditions $M_{11} = M_{TT} = 1+\lambda$, $M_{22} = M_{T-1,T-1} = 1+5\lambda$ and $M_{12} = M_{21} = M_{T,T-1} = M_{T-1,T} = -2\lambda$ and zeros elsewhere.

The transfer function of the HP filter has a smooth transition from the stopband to the passband. An important consequence of this gradual ascend is that a considerably large portion of low frequency components is passed through the highpass filter, a phenomenon that is especially pronounced when the underlying data series is integrated. In this case the HP filter generates strong cycles that lie to the left of the desired passband, similarly to case of approximate bandpass filters. This issue is well-known, see for example Cogley and Nason (1995), Guay and St-Amant (1996) and Pedersen (2001) for a detailed discussion.

A question that has received little attention so far is whether it is meaningful to compare, like Pederson does, the performance of the asymptotic HP filter to the ideal bandpass filter. If the ideal bandpass filter serves as the benchmark, the optimal approximation falls into the class of filters discussed in the previous section (3.1). In this section we therefore assume instead that the asymptotic HP filter, as described in equation (32) is the benchmark and emphasize the fact that the standard implementation using equation (34) is not the best finite sample approximation.

As an example, consider figure (4), where the HP filter is applied to a random walk with 29 observations. The coefficients of the optimal finite sample filter are obtained by the methodology described in section (2.4). The raw coefficients B_j are found by numerical integration of the inverse Fourier transform[5]

$$B_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{4\lambda[1-cos(\omega)]^2}{4\lambda[1-cos(\omega)]^2+1} e^{i\omega j} d\omega. \tag{35}$$

The optimal finite sample filter is then implemented according to the rules of propositions (2) and (3). In the middle of the sample both the standard filter as well as the optimal filter approximate the ideal filter very well. Since both filters are symmetric, there is no phase shift. However, at the end of the sample the standard HP approximation exaggerates the gain and experiences a large phase shift (up to π radians) at low frequencies. In contrast the optimal finite sample filter underestimates the gain, but has a phase shift that is only half the size of that of the standard HP implementation.

3.3 The Butterworth filter

Recently[6], Pollock (2000) and Trimbur and Harvey (2001) propose the use of a digital Butterworth filter as an approximation to a square wave filter to detrend economic time-series. The Butterworth filter, which is familiar in electrical engineering and audio-acoustic signal processing, is characterized by a gain function that is maximally flat (in the sense of the best Taylor approximation) in the passband and monotone between pass- and stopbands. This monotonicity comes at the price of a decrease in steepness in the transfer function, as compared to other classic IIR filters or the approximate bandpass filter discussed in section 3.1.

[5]This problem is well suited for numerical integration, since the integrand is a smooth function. At larger sample sizes Simpson's rule quadrature produced artifacts at higher coefficients, a problem which was avoided by switching to a higher order method such as adaptive Lobatto quadrature.

[6]An earlier application of the Butterworth filter for detrending economic time series can be found in Stock and Watson (1990).

The digital version of the Butterworth highpass filter is described by the rational polynomial expression (the filter's z-transform)

$$\psi_H(z) = \frac{\lambda(1-z)^n(1-z^{-1})^n}{(1+z)^n(1+z^{-1})^n + \lambda(1-z)^n(1-z^{-1})^n}. \tag{36}$$

Its time-domain representation (impulse response sequence) can be obtained by substituting z for the lag-operator L, while substituting z for $e^{-i\omega}$ gives the frequency-response function. The parameter n is referred to as the order of the filter and determines the steepness of the ascend between the stopband and the passband. The parameter λ determines the cutoff frequency ω_c such that for the highpass filter described above

$$\lambda = \left[\frac{1}{\tan\frac{\omega_c}{2}}\right]^{2n}, \tag{37}$$

$$\psi_H(e^{-i\omega})|_{n\to\infty} = \begin{cases} 1, & \text{if } \omega > \omega_c \\ 0, & \text{otherwise.} \end{cases} \tag{38}$$

Therefore, while the Butterworth filter shares some similarities with the Hodrick-Prescott filter (monotonicity and flatness), it is more flexible due the fact that the passband and steepness are controlled by two different parameters (for the HP filter both properties are controlled by λ). The order of the filter also equals the number of poles and zeros of the Laurent series represented by equation (36). The filter is stable as long as the poles remain inside the complex unit disk, however, increasing n moves the modulus of the poles closer to one. An increase in steepness of ascend therefore comes at the price of a decrease in stability.

Due to the recursive nature of the Butterworth filter problems arise for short and nonstationary data series. The main difficulty is to provide plausible initial and end conditions. A common approach is to extend the sample by forecasting and backcasting, however a bad choice of starting values can affect the entire sample (a problem known as the 'transient effect').

Pollock derives a specialized finite-sample version of the Butterworth filter on the basis of signal extraction theory. This approach is optimal if the data is consistent with the statistical model upon which the filter is based. The model under consideration is given by

$$\begin{aligned} y_t &= s_t + c_t \\ &= \frac{(1+L)^n}{(1-L)^d}\nu_t + (1-L)^{n-d}\epsilon_t, \quad \nu_t \sim N(0,\sigma_\nu^2), \quad \epsilon_t \sim N(0,\sigma_\epsilon^2), \end{aligned}$$

where s_t is the trend component extracted by the Butterworth lowpass filter and c_t is the cyclical component extracted by the Butterworth highpass filter. For a sample with N observations, the approximation is then given by[7]

$$\hat{c} = (-1)^d \lambda \mathbf{\Sigma} \mathbf{Q}(\mathbf{\Omega}_L + \lambda \mathbf{\Omega}_H)^{-1}\mathbf{Q}' y, \tag{39}$$

where $\mathbf{\Sigma}$ is an $N \times N$ Toeplitz matrix generated by $(1-z)^{n-d}(1-z^{-1})^{n-d}$, $\mathbf{\Omega}_L$ and $\mathbf{\Omega}_H$ are $(N-d)\times(N-d)$ Toeplitz matrices generated by $(1+z)^n(1+z^{-1})^n$ and $(1-z)^n(1-z^{-1})^n$, respectively, and $\mathbf{Q}$ is an $N \times (N-d)$ matrix with the coefficients of the polynomial $(1-L)^d$ in the elements with index (j,k), $j = 1, ..., N$ and $k = j, ..., j+d$. $\mathbf{Q}$ effectively operates as a d-fold differencing device.

However, as Pollock (1999, p. 607) notes, filters are usually selected not for their conformity with a specific data generating process, but rather with a view to their

[7] The actual solution described in Pollock (2000) is

$$\hat{c} = \lambda \mathbf{\Sigma} \mathbf{Q}(\mathbf{\Omega}_L + \lambda \mathbf{\Omega}_H)^{-1}\mathbf{Q}' y,$$

and does not include the factor $(-1)^d$.

frequency-response characteristics. In accordance with the methodology discussed in section 2.4 the optimal finite sample filter for a random walk model is constructed by first computing the raw coefficients via numerical integration of the inverse Fourier transform and then applying the rules from propositions 2 and 3.

Figure 5 compares the performance of the filter implied by equation (39) (*Pollock*) and the Kolmogorov-Wiener approximation of the Butterworth filter (*I(1) filter*) for an integrated process with sample size 29. The parameters are $n = 8$, $\omega_c = \frac{3\pi}{8}$, and $d = 1$ (since the process is assumed to be $I(1)$. It is remarkable that both filters provide an almost perfect approximation to the asymptotic filter in the middle of the sample. This may be explained by the fact that the filter coefficients of the asymptotic filter die out fairly quickly. However, at the end of the sample, the filter, as implemented by Pollock, shows a rather strong deviation from its ideal transfer function, both in terms of the gain function as well as the phase shift.

4 Application: End of sample filters for current analysis

In the previous section it was demonstrated that finite sample approximations of some well-known infinite impulse response (IIR) filters provide a good fit in terms of frequency response performance in the middle of the sample, but show sizable discrepancies towards the beginning and the end of the sample. It was also shown that these discrepancies can be mitigated to a significant extent by using the mean-squared error minimizing formulas introduced in section 2. A question that arises naturally is how these facts affect the performance of mechanical end-of-sample filters, which have been frequently used by central banks and other research institutions for current analysis purposes. This issue has been recently addressed by van Norden (2002), who uses an instance of the optimal bandpass filter, discussed in section 3.1, such that the end-of-sample estimate of the filtered component y_T is given by

$$\hat{y_T} = \sum_{j=-(T-1)}^{0} \hat{B}_j x_{T+j}. \tag{40}$$

$\{\hat{B}_j\}_{j=-(T-1)}^{0}$ minimizes

$$\mathrm{E}[y_T - \hat{y}_T | \{x_t\}_{t=-\infty}^{\infty}]^2, \tag{41}$$

where y_T is the output of a hypothetical ideal square-wave filter. In this section we compare the end-of-sample performance of the three filters discussed in the previous section (Bandpass filter, Hodrick-Prescott filter and Butterworth filter). The data sample used is the logarithm of quarterly real U.S. GDP from 1946:1 to 2000:1 made available electronically by the Federal Reserve Bank of St. Louis (shown in figure 6).

The sample of 220 observations was then truncated on each side by 50 observations, to allow for a startup sequence for the end-of-sample filters and the construction of a 2 sided mid-sample filter with a window-size of 101 observations that serves as a benchmark. In the previous section it was shown that for all three filters under consideration a symmetric truncation provides a very good fit when the sample size is 29, hence in this case the mid-sample filter may be regarded as a very close approximation to the asymptotic filter. The past information for the end-of-sample filters starts with a back-lag of 50 observations in 1958:3 and increases to 170 observations in 1987:3.

For the bandpass filter the passband is set for the interval $[\frac{\pi}{16}, \frac{\pi}{3})$, which coincides with the Burns-Mitchell (1946) taxonomy of business cycles with a wavelength between 8 and 32 quarters. The smoothing parameter of the Hodrick-Prescott filter is set to the usual value of $\lambda = 1600$. Finally, the cutoff frequency of the Butterworth filter is set to $\omega_c = \frac{\pi}{16}$ to allow a comparison with the approximate bandpass filter and the order, which determines the steepness of the ascend between stopband and passband, is set to $n = 8$.

It is widely agreed among economists that GDP is a described by a difference stationary data generating process (see for example the seminal work of Nelson and Plosser, 1982). For simplicity we assume that a simple random walk with a drift is a good approximation, so that GDP follows a pure random walk after removing a constant and the time trend. Therefore the optimal end-of-sample filter is constructed by applying the rules from propositions 2 and 3 to the infinite impulse response sequences of the ideal filters. This procedure is consistent with the findings of Christiano and Fitzgerald (1999) who show that accounting for the random walk component outweighs the importance of adjustments for serial correlation in the stationary component.

In figure 7 the optimal end-of-sample filters are compared to their standard finite sample counterparts (Dirichlet-truncation for the bandpass filter, the solution to the minimization problem (33) for the HP filter and Pollock's (2000) model based implementation for the Butterworth filter) as well as the benchmark set by the mid-sample filter. With the exception of Pollock's finite sample implementation of the Butterworth filter, all cyclical components are qualitatively similar and have downturns that coincide with NBER-type recessions. Summary statistics of the filtered series, shown in table 1, are consistent with the transfer-function comparisons of the previous section. In particular, the output of the optimal bandpass filter is relatively close to the one of the standard implementation, but has lower variance since fewer low frequency components are passed through. For the HP filter the gain of the optimal filter is only half of that of the ideal filter and the standard implementation, which translates into a lower variance of the filtered series. The main benefit of the optimal filter in this case is the fact that the phase effect is cut by half at the low frequencies. The biggest difference between the end-of-sample filters occurs in the case of the Butterworth filter, where the finite sample filter as implemented by Pollock has an artificial peak in its gain function around $\frac{\pi}{3}$ gradients (see figure 5), with the consequence that the filtered series contains a considerable proportion of artificial low-frequency components and its variance is seven times larger than that of the series generated by the mid-sample filter. The poor performance of Pollock's finite sample implementation of the Butterworth filter may be partly a result that for low cutoff frequencies the expression $\boldsymbol{\Omega}_L + \lambda \boldsymbol{\Omega}_H$ in equation (39) becomes nearly singular. This leads to inaccurate numerical results. An interesting observation is that the correlation between the optimal filter and the mid-sample filter σ_{om} is larger than the correlation between the standard filter and the mid-sample filter σ_{sm} in the cases of the HP filter and the Butterworth filter, but not for the bandpass filter. It can be confirmed, however, that the optimal filter dominates the standard filter in terms of the mean squared error criterion, by constructing the statistics

$$I_j \equiv \frac{1}{T-1} \sum_{t=1}^{T} (y_{j,t} - y_{m,j})^2 \tag{42}$$

and

$$R_j \equiv \frac{I_j}{\sigma_m^2}. \tag{43}$$

I_s and I_o are the mean squared deviations of the series generated by the standard and optimal filter, respectively, from the series generated by the mid-sample filter. R_s and R_o are standardized by the variance of the mid-sample series. A comparison of these statistics indicates that there are small, but noticeable improvements in the cases of the bandpass filter and the Hodrick-Prescott filter and a very large gain in accuracy in the case of the Butterworth filter.

A resampling method, based on the recursive bootstrap (see e.g. Freedman and Peters, 1984) is used to draw inference on the sample statistics. As opposed to the moving block bootstrap developed by Künsch (1989) and the stationary bootstrap proposed by Politis and Romano (1994), the recursive bootstrap approaches the problem of dependence in the data by fitting a parametric time series model. For the U.S. output data an ARIMA(1,1,0) model was found to effectively pre-whiten the data[8] and the residuals were then used to generate the bootstrap resamples in a recursive way. The bootstrap statistics indicate that the confidence intervals of the sample statistics (correlation and squared deviation from the ideal filter) are large. However, the confidence intervals of the absolute and relative squared deviation are small for the optimal filter, compared to the standard versions. The means of the bootstrap distribution indicates that the optimal filter consistently outperforms standard filters. The only exception occurs for the relative squared deviation of the Hodrick-Prescott filter. In this case the distribution of the standard implementation has virtually the same mean as the distribution of the optimal filter.

5 Conclusions

This paper discusses how linear filters with long and possible infinite impulse response sequences can be implemented for finite time-series in an optimal fashion.

[8]This is a commonly used parametric form for the U.S. output series, see, for example Watson (1986).

The analysis is based on Hilbert space projections and the optimization problem can be formulated either in the time- or in the frequency domain. The latter representation lends itself to the intuitive interpretation that the spectral density acts as a weighing function for the accuracy of the finite sample approximation.

For stationary time-series the optimal finite sample filter coefficients are a linear combination of the coefficients of the ideal filter sequence, with the autocovariance sequence coefficients acting as weights. This closed form solution is consistent with the above mentioned frequency domain interpretation, since the autocovariance function and the spectral density are a Fourier pair. For integrated time-series a necessary condition for optimality is that the sum of all finite sample filter coefficients is equal to the gain of the ideal filter at zero frequency. This result derives from the fact that the (pseudo) spectrum of an integrated process has infinite power at the zero frequency and any discrepancy of the filter approximation would lead to an infinite value of the objective function, and an ill-defined optimization problem. For pure random walks the adjustments affect only the first and last coefficient of the sample, for symmetric filters these adjustments can be expressed in a finite sum even if the impulse response sequence is infinite. For ARIMA models an optimal finite sample approximation is constructed by combining the results of propositions 1 to 3.

The optimal filter can be implemented by of first estimating the autocovariance function of the stationary component of the time-series (either directly, or by fitting an ARIMA model). In the case where no analytic expression of the impulse response function of the ideal filter exists, the filter coefficients can be computed by numerical integration of the inverse Fourier transformation of the transfer function.

An empirical example shows that the proposed solutions improve the accuracy of end-of-sample approximations for the bandpass filter, the Hodrick-Prescott filter and the Butterworth filter. Apart from these three filters the methodology of this paper is relevant for a wide range of filters with long impulse response sequences that fall outside the sample. In particular this group contains all filters with a rational lag polynomial, such as the Beveridge-Nelson smoother that was recently proposed by Proietti and Harvey (2001).

Matlab code for the examples and figures in the paper is available from the author.

Appendix

A Proof of proposition 1

An absolutely summable (possibly complex) function $\gamma(\cdot)$ defined on the integers is the autocovariance function of a stationary process if (and only if)

$$f_x(\omega) = \sum_{k=-\infty}^{\infty} e^{ik\omega}\gamma_k \geq 0, \quad \omega \in [-\pi, \pi],$$

in which case $f_x(\omega)$ is the spectral density of $\gamma(\cdot)$[9].

The optimization criterion can then be rewritten as

$$\begin{aligned} V(\hat{B}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |B(e^{i\omega}) - \hat{B}(e^{i\omega})|^2 f_x(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} |B(e^{i\omega}) - \hat{B}(e^{i\omega})|^2 e^{ik\omega}\gamma_k d\omega \\ &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma_k \int_{-\pi}^{\pi} |B(e^{i\omega}) - \hat{B}(e^{i\omega})|^2 e^{ik\omega} d\omega, \end{aligned}$$

where the interchange of summation and integration is justified by the Fubini-Tonelli theorem since $\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k=-\infty}^{\infty} |B(e^{i\omega}) - \hat{B}(e^{i\omega})|^2 |e^{ik\omega}\gamma_k| d\omega < \infty$ by square summability of $\{B_j\}$ and absolute summability of $\gamma(\cdot)$.

The first-order conditions with respect to $\hat{B}_j$ are

$$\sum_{k=-\infty}^{\infty} \gamma_k \int_{-\pi}^{\pi} \hat{B}(e^{i\omega}) e^{i\omega(j-k)} d\omega = \sum_{k=-\infty}^{\infty} \gamma_k \int_{-\pi}^{\pi} B(e^{i\omega}) e^{i\omega(j-k)} d\omega.$$

---

[9] This definition differs slightly from the usual definition, which relates the covariance function and the spectral density as the the Fourier pair

$$f_x(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{-ik\omega}\gamma(k) \quad \Leftrightarrow \quad \gamma(k) = \int_{-\pi}^{\pi} e^{ik\omega} f_x(\omega) d\omega,$$

(see e.g. Brockwell&Davis (1991), p. 120).

Since $\{e^{i\omega n}\}$ are orthonormal, all summands with a non-zero exponent disappear and the first-order conditions simplify to

$$\hat{B}\gamma_0 + \sum_{k=1}^{\infty}[\hat{B}_{j+k} + \hat{B}_{j-k}]\gamma_k = B_j\gamma_0 + \sum_{k=1}^{\infty}[B_{j+k} + B_{j-k}]\gamma_k, \quad j = -n_1, ..., n_2$$

or

$$\hat{B}_j\gamma_0 + \sum_{k=1}^{\min(Q,n_1+j)} \hat{B}_{j-k}\gamma_k + \sum_{k=1}^{\min(Q,n_2-j)} \hat{B}_{j+k}\gamma_k = B_j\gamma_0 + \sum_{k=1}^{Q}[B_{j+k}+B_{j-k}]\gamma_k, \quad j = -n_1, ..., n_2$$

if $\gamma_k = 0$ for $k \geq Q$.

In matrix notation this system of equations can be written as

$$\mathbf{\Gamma}B = \hat{\mathbf{\Gamma}}\hat{B}$$

where $\mathbf{\Gamma}, \hat{\mathbf{\Gamma}}, B$ and $\hat{B}$ are described in proposition 1. $\hat{\mathbf{\Gamma}}$ is a symmetric matrix with nonzero diagonal entries and, in general, invertible.

$\square$

B Proof of proposition 2

The criterion of the optimization problem is given by

$$\begin{aligned} V(\hat{B}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}|B(e^{i\omega}) - \hat{B}(e^{i\omega})|^2\frac{\sigma^2}{|1-e^{i\omega}|^2}d\omega \\ &= \frac{\sigma^2}{2\pi}\|\frac{B(e^{i\omega}) - \hat{B}(e^{i\omega})}{1-e^{i\omega}}\|^2. \end{aligned}$$

Since for any function $f \in L^2$, $|\int f^2| \leq \int |f|^2$ (*Cauchy-Schwarz*), a necessary condition for the criterion to be finite is that

$$\left|\int_{-\pi}^{\pi}\left(\frac{B(e^{i\omega}) - \hat{B}(e^{i\omega})}{1-e^{i\omega}}\right)^2 d\omega\right|$$

$$= \left| \oint_{|z|=1} \frac{(B(z) - \hat{B}(z))^2}{iz(1-z)^2} dz \right| < \infty.$$

The denominator has a double pole at $z = 1$ and its contour integral is unbounded

$$\left| \oint_{|z|=1} \frac{1}{z(1-z)^2} dz \right| = \infty,$$

which makes it necessary to impose restrictions on the numerator to ensure finiteness. Since $B(z)$ and $\hat{B}(z)$ are polynomials, the only possible restriction is of the form

$$B(z) - \hat{B}(z) = g(z)[C(z) - \hat{C}(z)],$$

where $g(z)$, $C(z)$ and $\hat{C}(z)$ are also polynomials. The lowest order polynomial which removes the singularity is $g(z) = 1 - z$, that is, we restrict $B(z) - \hat{B}(z)$ to have a zero at $z = 1$, offsetting the pole at the same location. Since 1 is a root of $B(z) - \hat{B}(z)$, the transfer functions of both filters have to be equal to some constant β at the zero frequency

$$B(1) = \sum_{k=-\infty}^{\infty} B_j = \sum_{k=-\infty}^{\infty} \hat{B}_j = \hat{B}(z) \equiv \beta.$$

To solve for the coefficients of $C(z)$ and $\hat{C}(z)$, define $a = C(z) - \hat{C}(z)$. Then

$$\begin{aligned} B(z) - \hat{B}(z) &= \sum_{j=-\infty}^{\infty} (B_j - \hat{B}_j) z^j \\ &= (1-z) \sum_{j=-\infty}^{\infty} a_j z^j \\ &= \sum_{j=-\infty}^{\infty} (a_j - a_{j-1}) z^j. \end{aligned}$$

Therefore $B_j - \hat{B}_j = a_j - a_{j-1}\ \forall j \in \mathbb{Z}$. There are three different cases to consider:

- $j < -n_1$: $\hat{B}_j = 0$, therefore $a_j = a_{j-1} + B_j$ and

$$a_j = \sum_{k=-\infty}^{j} B_k.$$

- $-n_1 \leq j \leq n_2$: $a_j = a_{j-1} + B_j - \hat{B}_j$, therefore

$$a_j = \sum_{k=-\infty}^{j} B_k - \sum_{k=-n_1}^{n_2} \hat{B}_k.$$

- $n_2 < j$: $\hat{B}_j = 0$, therefore $a_j = a_{j-1} + B_j$ and

$$a_j = \sum_{k=-\infty}^{j} B_k - \sum_{k=-n_1}^{n_2} \hat{B}_k = \sum_{k=-\infty}^{j} B_k - \beta = -\sum_{k=j+1}^{\infty} B_j.$$

We can therefore describe the polynomials $C(z)$ and $\hat{C}(z)$ as

$$C_j = \begin{cases} \sum_{k=-\infty}^{j} B_k & \text{if } j \leq n_2 \\ -\sum_{k=j+1}^{\infty} B_k & \text{if } j > n_2 \end{cases}, \qquad \hat{C}_j = \sum_{k=-n_1}^{n_2} \hat{B}_j.$$

Note that while $\hat{B}(z)$ is a polynomial from $-n_1$ to n_2, $\hat{C}(z)$ has exponents from $-n_1$ to $n_2 - 1$, reflecting the loss of flexibility due to the integration-restriction.

It is now possible to restate the optimization problem in terms of C and $\hat{C}$ as

$$V(\{\hat{C}_j\}_{j=-n_1}^{n_2-1}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |C(e^{i\omega}) - \hat{C}(e^{i\omega})|^2 d\omega.$$

A sufficient condition for the integral to be finite is that $\sum_{j=-\infty}^{\infty} |C_j|^2 < \infty$. From the above definition of C_j and keeping in mind that finite sums of finite elements are bounded and can therefore be omitted, the criterion can be restated as

$$\sum_{j=-\infty}^{0} |\sum_{k=-\infty}^{j} B_k|^2 + \sum_{j=1}^{\infty} |\sum_{k=j+1}^{\infty} B_k|^2 < \infty.$$

The optimization problem $V(\hat{C})$ is identical to the the problem $V(\hat{B})$ for white noise series in appendix A, with first order conditions

$$\hat{C}_j = C_j, \quad j = -n_1, ..., n_2 - 1.$$

Together with the constraint

$$\sum_{j=-n_1}^{n_2} \hat{B}_j = \beta,$$

the solution forms a system of N linear equations in N unknowns, with matrix representation

$$\left[\begin{array}{c} \mathbf{D} \\ \iota \end{array} \right] \hat{B} = \left[\begin{array}{c} C \\ \beta \end{array} \right].$$

□

C Proof of proposition 3

With symmetric filters ($B_j = B_{-j}$) it is possible to express the (semi)-infinite sums of proposition 2 with finite closed forms that can be conveniently expressed in matrix notation.

As before we define $B(1) = \sum_{k=-\infty}^{\infty} B_k = \beta$. Then for $j < 0$:

$$\begin{aligned} C_j &= \sum_{k=-\infty}^{j} B_k = \sum_{-\infty}^{\infty} B_k - \sum_{-j}^{\infty} B_k - \sum_{j+1}^{-j-1} B_k \\ 2\sum_{k=-\infty}^{j} B_k &= \beta - B_0 - 2\sum_{k=1}^{|j|-1} B_k \\ C_j &= \frac{\beta}{2} - \frac{B_0}{2} - \sum_{k=1}^{|j|-1} B_k \end{aligned}$$

Similarly, for $0 \leq j \leq n_2$:

$$\begin{aligned} C_j &= \sum_{k=-\infty}^{j} B_k = \beta - \sum_{k=j+1}^{\infty} B_k \\ &= \beta - C_{-j} \end{aligned}$$

$$
\begin{aligned}
&= \beta - \frac{\beta}{2} + \frac{B_0}{2} + \sum_{k=1}^{j} B_k \\
&= \frac{\beta}{2} + \frac{B_0}{2} + \sum_{k=1}^{j} B_k.
\end{aligned}
$$

In matrix notation (using the definitions from proposition 2) this can be written as

$$C = \mathbf{M}B + \frac{\beta}{2}\tau.$$

□

References

Baxter, M., King, R. 1999. Measuring business-cycles: approximate band-pass filters for economic time series. *Review of Economics and Statistics* **81**(4): 575–593.

Burns, Arthur M., Mitchell, Wesley C. 1946 *Measuring Business Cycles.* New York, N.Y.: NBER.

Christiano, Lawrence J., Fitzgerald, Terry J. 1999. The bandpass filter. working paper 99/06, Federal Reserve Bank of Cleveland.

Cogley, Timothy, Nason, James N. 1995. Effects of the Hodrick-Prescott filter on trend and difference stationary time series, implications for business cycel research. *Journal of Economic Dynamics and Control* **19**: 253–278.

Freedman, D.A., Peters, S.C. 1984. Bootstrapping a regression equation: Some empirical results. *Journal of the American Statistical Association* **79**: 97–106.

Gibbs, Josiah Willard 1899. Fourier's series: A letter to the editor. *Nature* **59**: 606.

Granger, Clive 1966. The typical shape of an economic variable. *Econometrica* **34**(1): 150–161.

Guay, Alain, St-Amant, Pierre 1996. Do mechanical filters provide a good approximation of business cycles. Technical Report 78, Bank of Canada.

Hodrick, Robert, Prescott, Edward 1997. Post-war business cycles: An empirical investigation. *Journal of Money, Credit and Banking* **29**(1): 1–16.

Khintchine, A. 1934. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen* **109**: 604–615.

King, R.G., Rebelo, S.T. 1993. Low frequency filtering and business cycles. *Journal of Economic Dynamics & Control* **17**: 207–232.

Kolmogorov, A.N. 1941. Interpolation and extrapolation. *Bulletin de l'Academie des Sciences de U.S.S.R., Series Mathematics* **5**: 3–14.

Künsch, H.R. 1989. The bootstrap and jacknife for general stationary observations. *The Annals of Statistics* **17**: 1217–1241.

Lucas, R.E., Jr. 1980. Two illustrations of the quantity theory of money. *American Economic Review* **70**(5): 1005–1014.

Nelson, C.R., Plosser, C.I. 1982. Trends and random walks in macroeconomic time series. *Journal of Monetary Economics* **10**: 129–162.

Pedersen, Thorben Mark 2001. The Hodrick-Prescott filter, the Slutzky effect, and the distortionary effect of filters. *Journal of Economic Dynamics & Control* **25**: 1082–1101.

Politis, D., Romano, J. 1994. The stationary bootstrap. *Journal of the American Statistical Association* **89**: 1303–1313.

Pollock, D.S.G. 1999 *A Handbook of Time-Series Analysis, Signal Processing and Dynamics.* London: Academic Press.

—— 2000. Trend estimation and de-trending via rational square-wave filters. *Journal of Econometrics* **99**: 317–334.

Proietti, Tommaso, Harvey, Andrew 2000. A Beveridge-Nelson smoother. *Economics Letters* **67**: 139–146.

Reinsch, C.H. 1976. Smoothing by spline functions. *Numerische Mathematik* **10**: 177–183.

Stock, J., Watson, M. 1990. Business cycle properties of selected US economic time series 1959-1988. Working Paper 3376, NBER.

Trimbur, Thomas M., Harvey, Andrew C. 2001. General model-based filters for extracting cycles and trends in economic time series. Cambridge University, Faculty of Economics and Politics.

van Norden, Simon 2002. Filtering for current analysis. Working Paper 2002-28, Bank of Canada.

Watson, M. W. 1986. Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* **18**: 49–75.

Wiener, Norbert 1930. Generalised harmonic analysis. *Acta Mathematica* **35**: 117–258.

—— 1949 *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* Cambridge, MA: MIT Press.

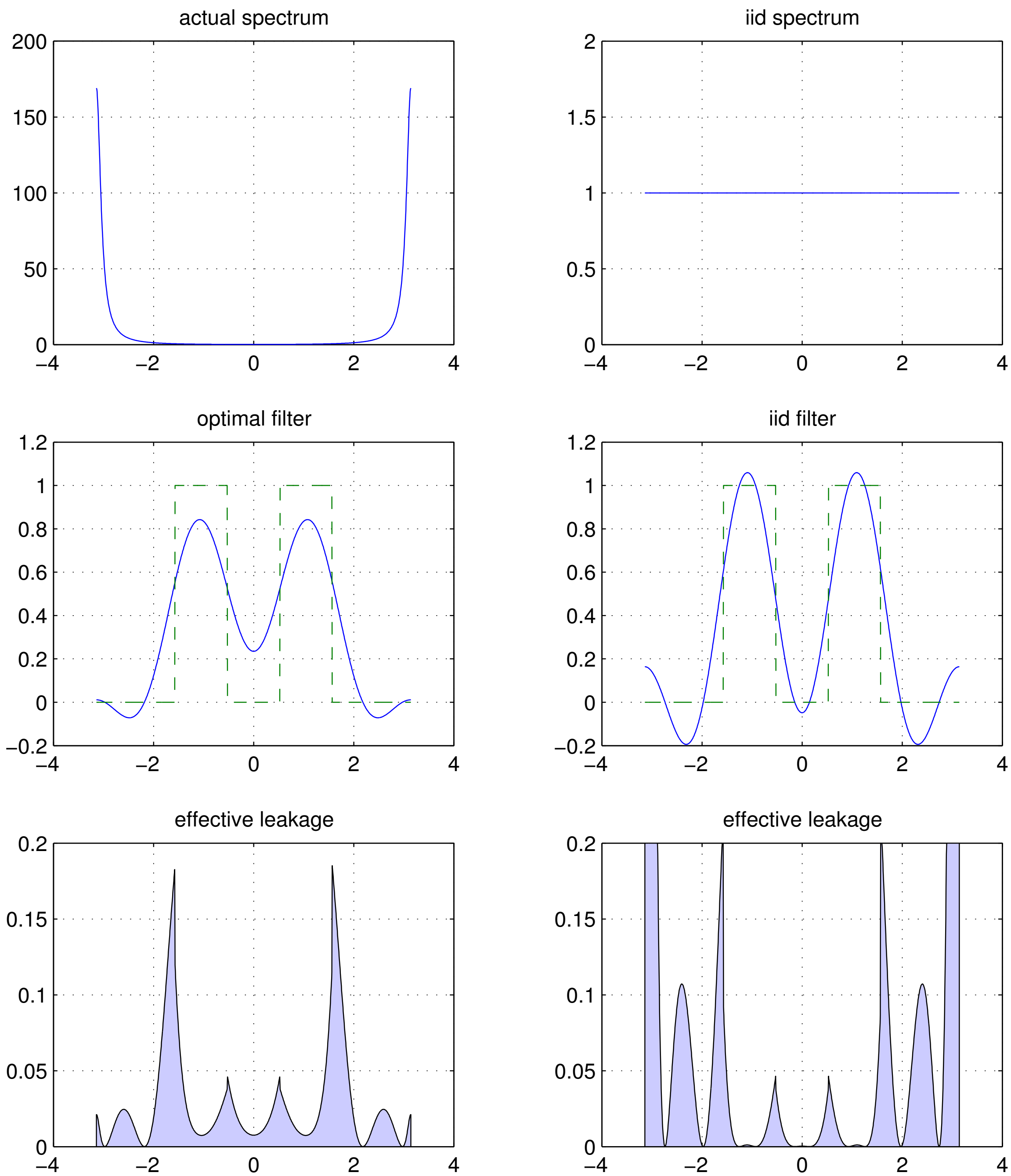

Figure 1: Filtering the ARMA(1,1) process of Example 1 for the passband $\pi/6, \pi/2]$ with 7 symmetric coefficients ($n_1 = n_2 = 3$). The left column shows the actual spectral density, the squared gain function of the filter and the squared leakage, weighted by the actual spectral density. The right column shows a standard *iid*-filter for comparison.

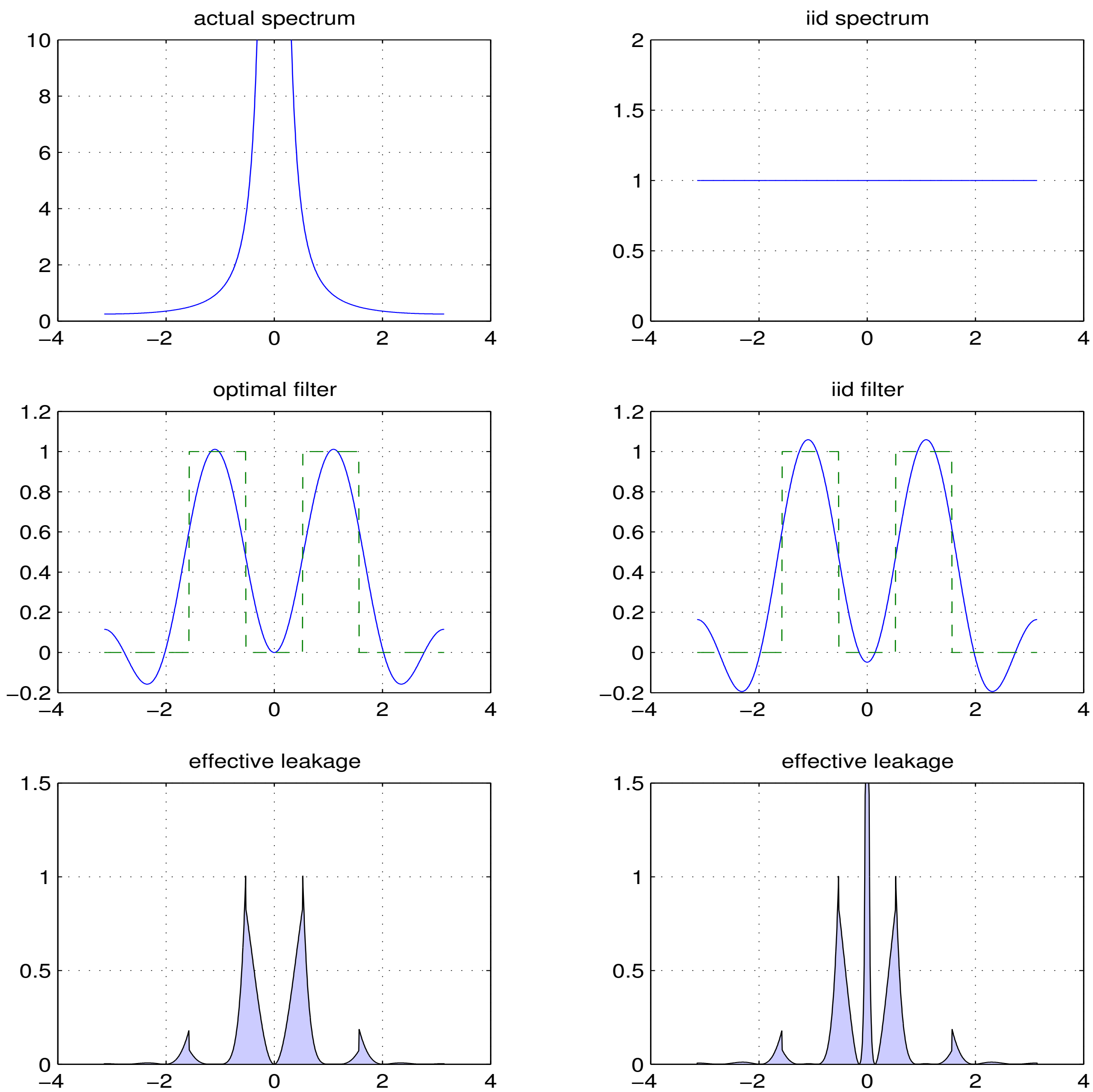

Figure 2: Transfer function of an optimal bandpass filter (with $a = \pi/6, b = \pi/2, n_1 = n_2 = 3$) for a random walk (*left*) and for *iid* disturbances (*right*). The bottom row shows effective leakage when the underlying process follows a random walk. For the *iid*-filter leakage approaches infinity at the origin.

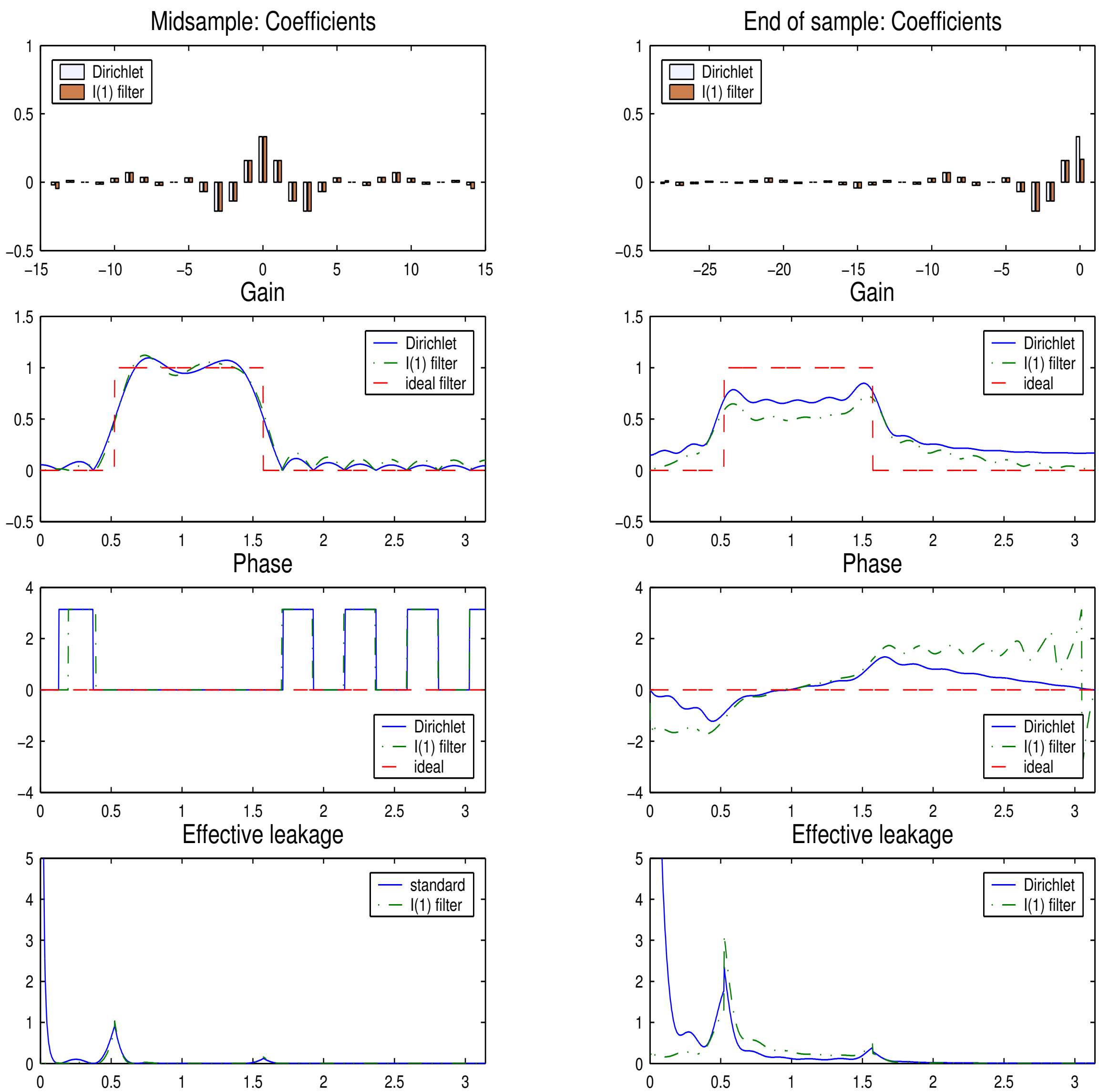

Figure 3: Comparison between the Dirichlet truncation and the optimal I(1) filter for a random walk with sample size 29. The filters are approximations for a bandpass filter with passband $[\frac{\pi}{6}, \frac{\pi}{2}]$.

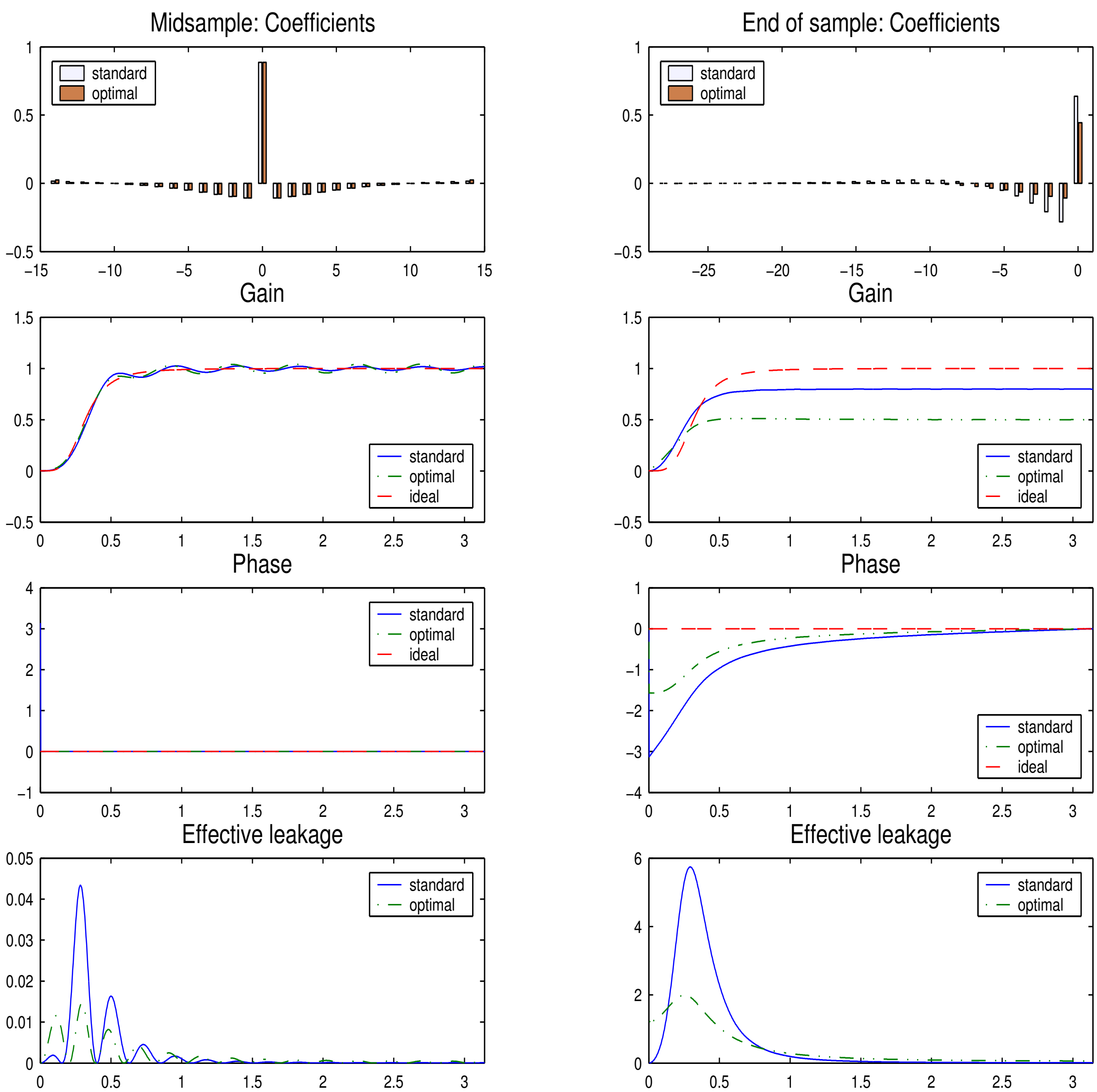

Figure 4: Implementation of the Hodrick-Prescott filter for a random walk process for a sample of 29 observations and $\lambda = 100$.

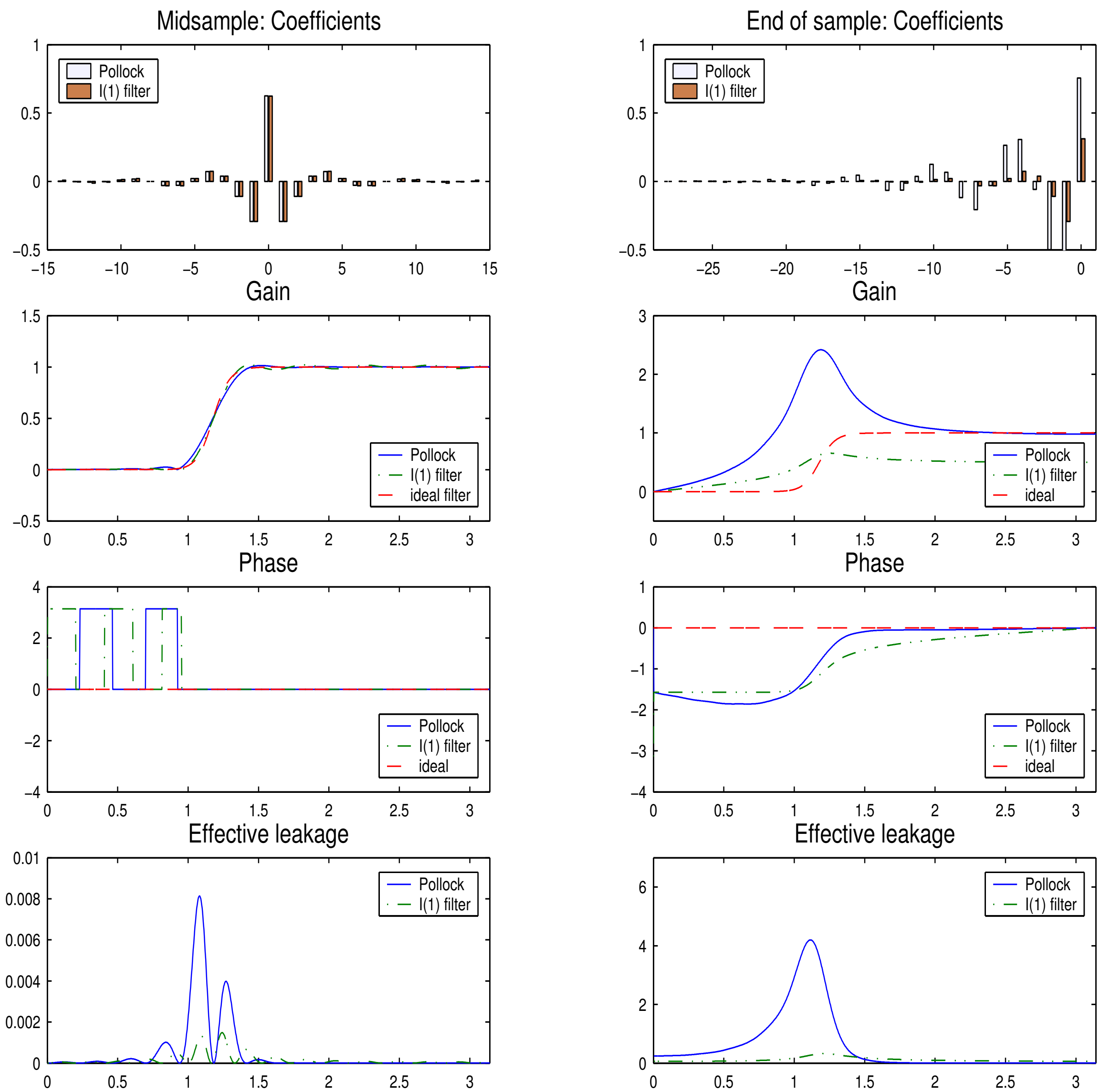

Figure 5: Implementation of the Butterworth filter for a random walk with sample size 29, cutoff frequency $\omega_c = \frac{3\pi}{8}$ and $n = 8$.

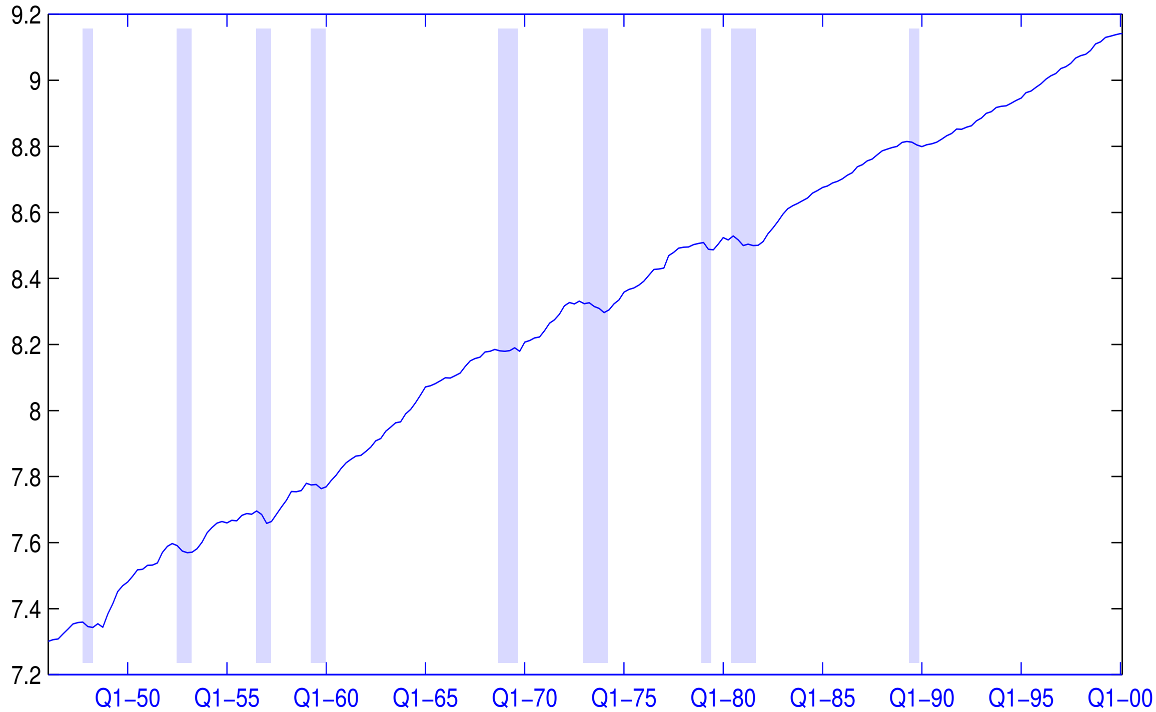

Figure 6: Log Quarterly U.S. GDP (1946-2000). NBER recessions are shaded.

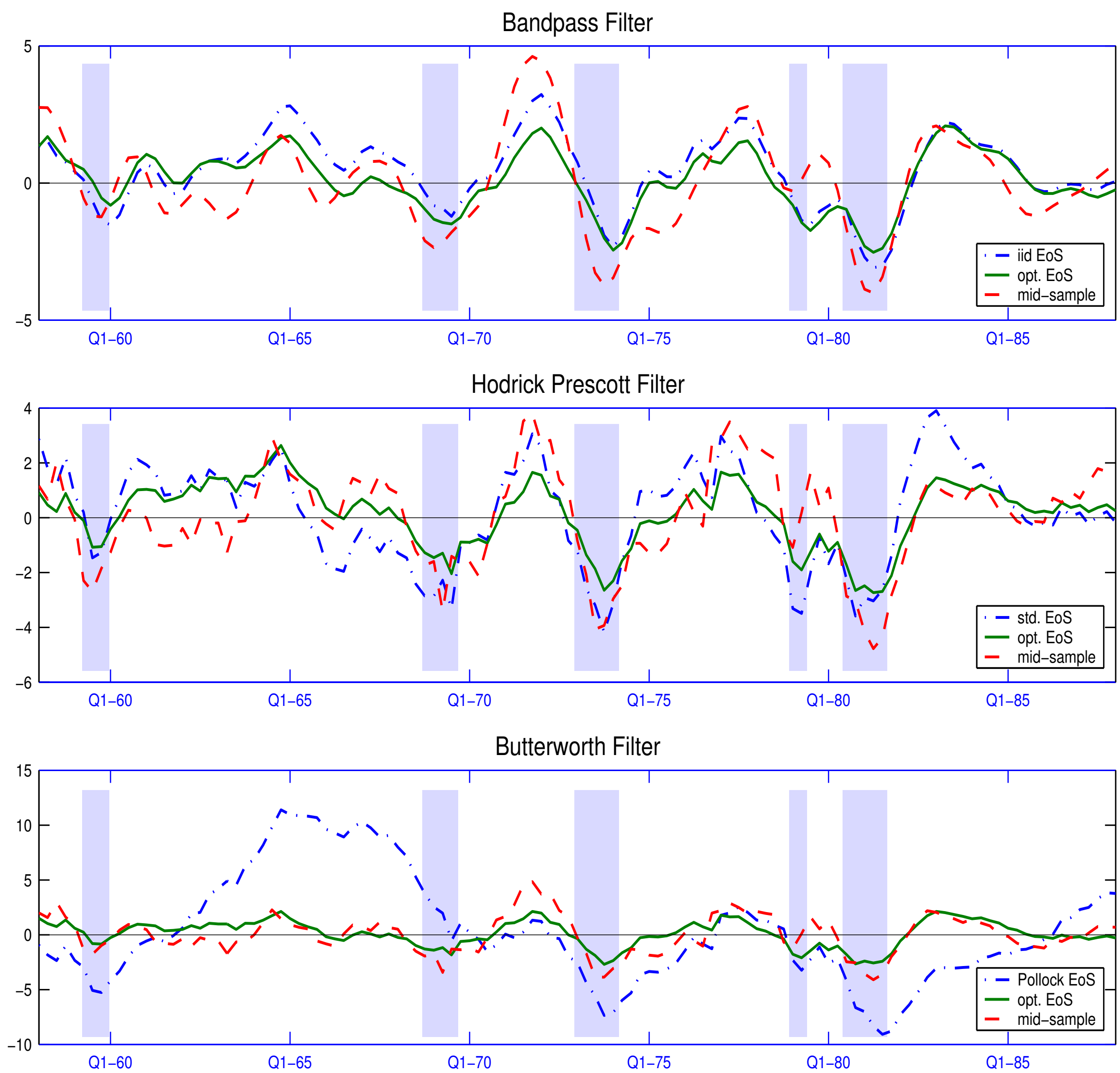

Figure 7: Comparison of end of sample filters. In each case the mid-sample filter provides a close approximation of the ideal filter. NBER recessions are shaded.

Table 1: Sample Statistics

Bandpass Filter

	Variance			Correlation	
σ_s^2	1.79	*[0.98, 2.01, 3.11]*	ρ_{sm}	0.79	*[0.48, 0.67, 0.84]*
σ_o^2	1.17	*[0.56, 1.23, 1.94]*	ρ_{om}	0.77	*[0.43, 0.65, 0.82]*
σ_m^2	3.06	*[1.16, 2.41, 3.80]*	ρ_{so}	0.92	*[0.72, 0.89, 0.97]*
	Av. Sq. Deviation			**Rel. Sq. Deviation**	
I_s	1.3141	*[0.76, 1.60, 2.90]*	R_s	0.4284	*[0.33, 0.72, 1.54]*
I_o	1.3055	*[0.73, 1.38, 2.34]*	R_o	0.4256	*[0.37, 0.60, 0.98]*

Hodrick-Prescott Filter

	Variance			Correlation	
σ_s^2	3.23	*[1.58, 3.18, 4.83]*	ρ_{sm}	0.60	*[0.34, 0.56, 0.74]*
σ_o^2	1.35	*[0.71, 1.49, 2.30]*	ρ_{om}	0.75	*[0.50, 0.67, 0.81]*
σ_m^2	3.12	*[1.43, 2.82, 4.18]*	ρ_{so}	0.84	*[0.69, 0.82, 0.90]*
	Av. Sq. Deviation			**Rel. Sq. Deviation**	
I_s	2.4818	*[1.11, 2.66, 5.35]*	R_s	0.7933	*[0.51, 0.95, 1.61]*
I_o	2.4000	*[1.83, 2.55, 3.59]*	R_o	0.7671	*[0.62, 0.96, 1.47]*

Butterworth Filter

	Variance			Correlation	
σ_s^2	23.84	*[9.49, 35.29, 67.27]*	ρ_{sm}	0.29	*[0.07, 0.25, 0.48]*
σ_o^2	1.35	*[0.63, 1.38, 2.13]*	ρ_{om}	0.74	*[0.47, 0.66, 0.81]*
σ_m^2	3.07	*[1.27, 2.52, 3.83]*	ρ_{so}	0.36	*[0.08, 0.35, 0.59]*
	Av. Sq. Deviation			**Rel. Sq. Deviation**	
I_s	21.8680	*[8.46, 34.36, 84.59]*	R_s	7.1126	*[2.98, 14.88, 40.41]*
I_o	1.3966	*[0.80, 1.44, 2.33]*	R_o	0.4542	*[0.36, 0.60, 0.97]*

---

s, o and m denote the standard, optimal and ideal filter, respectively. In parenthesis: [lower bound of 95% confidence interval, mean, upper bound of 95% confidence interval] from 999 bootstrap resamples.