TREND ESTIMATION AND DE-TRENDING VIA RATIONAL SQUARE-WAVE FILTERS

by D.S.G. Pollock

Queen Mary and Westfield College

University of London

This paper gives an account of some techniques of linear filtering which can be used for extracting trends from economic time series of limited duration and for generating de-trended series. A family of rational square-wave filters is described which enable designated frequency ranges to be selected or rejected. Their use is advocated in preference to other filters which are commonly used in quantitative economic analysis.

1. Introduction

Recently there has been a growing interest amongst economists in the techniques for detrending nonstationary times series and for representing their underlying trends. Much of this interest has been associated with the resurgence of business cycle analysis in the field of macroeconomics (see, for example, Hartley *et al.* [6]).

The technique which has been used predominantly in trend estimation and detrending is that of linear filtering; and there are some unresolved problems. This paper proposes to tackle these problems by adopting some of the perspectives of electrical and audio-acoustic signal processing. The endeavour to adapt the signal processing techniques to the characteristics of economic data series, which are liable to be strongly trended and of a strictly limited duration, leads to some innovations which are presented in this paper.

Underlying the use of filtering techniques in business cycle analysis is the notion that an economic time series can be represented as the sum of a set of statistically independent components each of which has its own characteristic spectral properties. If the frequency ranges of the components are completely disjoint, then it is possible to achieve a definitive separation of the time series into its components. If the frequency ranges of the components overlap, then it is still possible to achieve a tentative separation in which the various components take shares of the cyclical elements of the time series.

The trend of a time series is the component which comprises its noncyclical elements together with the cyclical elements of lowest frequency. If the trend is a well-defined entity, then it should be possible to specify a cut-off frequency which delimits the range of its cyclical elements. The ideal filter for isolating such a trend would possess a passband, which admits to the estimated trend all elements of frequencies less that the cut-off value, and a stopband, which impedes all elements of frequencies in excess of that value.

On the other hand, if the frequency range of the trend overlaps substantially with the ranges of the adjacent components, then one might be inclined to adopt a filter which shows only a gradual transition from the passband to the stopband. In that case, the nominal cut-off point of the filter becomes the mid point of the transition band and the rate of transition becomes a feature of the filter which needs to be adjusted in view of the extent of the overlap.

The Hodrick–Prescott (H–P) filter (see [4], [7] and [10]), which has been used widely by business-cycle analysts, possesses only a single adjustable parameter which affects both the rate of transition and the location of the nominal cut-off point. This so-called smoothing parameter is liable to be fixed by rule of thumb or by convention. The H–P filter is closely related to the Reinsch [14] smoothing spline which is used extensively in industrial design.

Whereas the H–P filter is an excellent device for representing the broad trend of a time series, it often fails in the more exacting task of generating a detrended series. In particular, it sometimes allows powerful low-frequency components to pass through into the detrended series when they ought to be impeded by the filter. This deficiency is the basis of the oft-repeated, albeit inaccurate, claim that the H–P filter is liable to induce spurious cycles in detrended data series.

An alternative approach to detrending, which has attracted economists, is a model-based approach which attempts to estimate the trend in the company of other identifiable components of the economic time series (see [8], [9] and [11], for examples). Such an approach avoids the use of a rule of thumb in determining the characteristics of the filter, but it imposes other features which may be undesirable. In particular, the model-based approach postulates a trend which is the product of an integrated moving-average process of a low order which has no firm upper limit to its frequency range.

In this paper, our primary objective is to design a filter with a well-defined cut-off point and with a rapid transition. The evidence of periodogram analysis suggest that, in many econometric time series, the trend component is clearly segregated from the remaining components; and this indicates a need for a filter which can effect a clear separation.

It transpires that the filter in question, which fulfils our design objectives, is an instance of the digital Butterworth filter which is familiar to electrical engineers (see, for example, Roberts and Mullis [15]). This filter is commonly regarded as the digital translation of an analogue design. In this light, it is interesting to discover that the filter arises independently in the digital domain in the pursuit of some simple design objectives.

The filter can also be derived by applying the Wiener–Kolmogorov theory of signal extraction to a specific statistical model. (See [11] and [16] for the original expositions of the theory.) We shall use a finite-sample version of Wiener–Kolmogorov approach in adapting the filter to cope with samples of limited duration. Our method of coping with the problem of non-stationarity will also be given a statistical justification.

2. Rational Square-Wave Filters

The filters which we shall consider in this paper are of the bidirectional variety which are applied by passing forwards and backward through the data series. The same filter is used in both directions. Thus, if $\{y(t); t = 0 \pm 1, \pm 2, \ldots\}$ represents a discrete-time signal, then the two filtering operations can be described by the equations

$$(1) \qquad \text{(i)} \quad \gamma(L)q(t) = \delta(L)y(t) \qquad \text{and} \qquad \text{(ii)} \quad \gamma(F)x(t) = \delta(F)q(t),$$

wherein $q(t)$ is the intermediate output from the forwards pass of the filter and $x(t)$ is the final output which is generated in the backward pass. Here $\gamma(L)$ and $\delta(L)$ stand for polynomials of the lag operator L, whilst $\gamma(F)$ and $\delta(F)$ have the inverse forwards-shift operator $F = L^{-1}$ in place of L.

The filter must fulfil the condition of stability which requires the roots of the polynomial equation $\gamma(z) = 0$ to lie outside the unit circle. Equivalently, the roots of $\gamma(z^{-1}) = 0$ must lie inside the unit circle.

For convenience, the two filters can be combined to form a symmetric two-sided rational filter

$$(2) \qquad \psi(L) = \frac{\delta(F)\delta(L)}{\gamma(F)\gamma(L)};$$

but, of necessity, this must be applied in the manner indicated by (1).

The effect of a linear filter is to modify a signal by altering the amplitudes of its cyclical components and by advancing and delaying them in time. These modifications are described respectively as the gain effect and the phase effect; and they are referred to jointly as the frequency response of the filter.

The frequency response of the filter is obtained by replacing the L by the complex argument z. The resulting complex-valued function can be expressed in polar form as

$$(3) \qquad \psi(z) = |\psi(z)| \exp[i\mathrm{Arg}\{\psi(z)\}],$$

where $|\psi(z)| = \sqrt{\psi(z)\psi(z^{-1})}$ is the complex modulus of $\psi(z)$, and $\mathrm{Arg}\{\psi(z)\} = \mathrm{Arctan}\{\psi^{\mathrm{im}}(z)/\psi^{\mathrm{re}}(z)\}$ is the argument of $\psi(z)$. The gain of the filter, which is a function of $\omega \in [0, \pi]$, is obtained by setting $z = e^{-i\omega}$ in $|\psi(z)|$. The phase lag of the filter, which is also regarded as a function of ω, is obtained by setting $z = e^{-i\omega}$ in $\mathrm{Arg}\{\psi(z)\}$.

The bidirectional rational filter of equation (2) fulfils the condition that $\psi(z) = \psi(z^{-1})$. This is the condition of so-called phase neutrality, and it implies that $\psi(z) = |\psi(z)|$ and that $\mathrm{Arg}(z) = 0$. In effect, the real-time phase lag which is induced by the forwards pass of the filter represented by equation

(1)(i) is exactly offset by a reverse-time phase lag induced by the backwards pass of (1)(ii).

In the terminology of digital signal processing, an ideal frequency-selective filter is a phase-neutral filter for which the gain is unity over a certain range of frequencies, described as the passband, and zero over the remaining frequencies, which constitute the stopband. In a lowpass filter ψ_L, the passband covers a frequency interval $[0, \omega_c)$ ranging from zero to a cut-off point. In the complementary highpass filter ψ_H, it is the stopband which stands on this interval. Thus

$$(4) \qquad |\psi_L(e^{i\omega})| = \begin{cases} 1, & \text{if } \omega < \omega_c \\ 0, & \text{if } \omega > \omega_c \end{cases} \qquad \text{and} \qquad |\psi_H(e^{i\omega})| = \begin{cases} 0, & \text{if } \omega < \omega_c \\ 1, & \text{if } \omega > \omega_c. \end{cases}$$

An ideal filter $\psi(z)$ fulfilling one or other of these conditions together with the condition of phase neutrality constitutes a positive semi-definite idempotent function. The function is idempotent is the sense that $\psi(z) = \psi^2(z)$. It is positive semi-definite in the sense that

$$(5) \qquad 0 \leq \frac{1}{2\pi i} \oint \lambda(z)\psi(z)\lambda(z^{-1}) \frac{dz}{z} = \frac{1}{2\pi i} \oint |\lambda(z)\psi(z)|^2 \frac{dz}{z},$$

where $\lambda(z)$ is any polynomial or power series in z. Here, the final equality depends upon the assumption of phase neutrality which, in combination with the condition of idempotency, implies that $\psi(z) = \psi^2(z) = \psi(z)\psi(z^{-1})$. When the locus of this contour integral is the unit circle, the expression stands for the integral of the gain of the composite filter $\lambda(\omega)\psi(\omega)$. Observe that an equality holds on the LHS when $\lambda(z) = 1 - \psi(z)$. In that case, $\lambda(\omega)$ represents an ideal filter which is the complement of $\psi(\omega)$ and which nullifies the output of the latter which is nonzero only over its passband.

The object in constructing a practical frequency-selective filter, is to find a rational function, embodying a limited number of coefficients, whose frequency response is a good approximation to the square wave. In fact, the idealised conditions of (4) are impossible to fulfil in practice. In particular, the stability condition affecting $\gamma(z)$, which implies that $\psi(z)$ must be a positive-definite function, precludes the fulfilment of the conditions. Moreover, improvements in the accuracy of the approximation, which are bound to be purchased at the cost of increasing the number of filter coefficients, are liable to exacerbate the problems of stability.

In this section, we shall derive a pair of complementary filters which fulfil the specifications of (4) approximately for a cut-off frequency of $\omega_c = \pi/2$. Once we have designed these prototype filters, we shall be able to apply a transformation which shifts the cut-off point from $\omega = \pi/2$ to any other point $\omega_c \in (0, \pi)$.

A preliminary step in designing a pair of complementary filters is to draw up a list of specifications which can be fulfilled in practice. We shall be guided by the following conditions:

$$
\begin{array}{lll}
\text{(i)} & \psi_L(z^{-1}) = \psi_L(z), \quad \psi_H(z^{-1}) = \psi_H(z), & \textit{Phase-Neutrality} \\
\text{(ii)} & \psi_L(z) + \psi_H(z) = 1, & \textit{Complementarity} \\
\text{(iii)} & \psi_L(-z) = \psi_H(z), \quad \psi_H(-z) = \psi_L(z), & \textit{Symmetry} \\
\text{(iv)} & |\psi_H(1)| = 0, \quad |\psi_H(-1)| = 1, & \textit{Highpass Conditions} \\
\text{(v)} & |\psi_L(1)| = 1, \quad |\psi_L(-1)| = 0. & \textit{Lowpass Conditions}
\end{array} \tag{6}
$$

As we have already noted, a bidirectional rational filter in the form of (2) fulfils already the condition (i) of phase neutrality. Given the specification of (2), the condition of complementarity under (ii) implies that the filters must take the form of

$$
\psi_L(z) = \frac{\delta_L(z)\delta_L(z^{-1})}{\gamma(z)\gamma(z^{-1})} \qquad \text{and} \qquad \psi_H(z) = \lambda\frac{\delta_H(z)\delta_H(z^{-1})}{\gamma(z)\gamma(z^{-1})}. \tag{7}
$$

where

$$
\gamma(z)\gamma(z^{-1}) = \delta_L(z)\delta_L(z^{-1}) + \lambda\delta_H(z)\delta_H(z^{-1}). \tag{8}
$$

Here λ is a so-called smoothing parameter which will be used for varying the cut-off point of the filter and which takes the value of $\lambda = 1$ in the prototype filters.

Next, the condition of symmetry under (iii) implies that, when it is reflected about the axis of $\omega = \pi/2$, the frequency response of the lowpass filter becomes the frequency response of the highpass filter. This implies that the cut-off point ω_c must be located at the mid-point frequency of $\pi/2$. The condition requires that $\lambda = 1$ and that $\gamma(z) = \gamma(-z)$, which implies that every root of $\gamma(z) = 0$ must be a purely imaginary number. The condition also requires that

$$
\delta_L(z) = \delta_H(-z) \qquad \text{and} \qquad \delta_H(z) = \delta_L(-z), \tag{9}
$$

which has immediate implications for equation (8).

It remains to fulfil the conditions (iv) and (v). Condition (iv) indicates that $\delta_H(z)$ must have a zero at $z = 1$, which is to say that it must incorporate a factor in the form of $(1 - z)^n$. Observe that, if the filter ψ_H is to be used in reducing an ARIMA(p, d, q) process to stationarity, then it is necessary that $n \geq d$ where d is the number of unit roots in the autoregressive operator. Condition (v) indicates that $\delta_L(z)$ must have a zero at $z = -1$, which is to say that it must incorporate a factor in the form of $(1+z)^n$. These conditions (iv) and (v) do not preclude the presence of further factors in $\delta_L(z)$ and $\delta_H(z)$. However, the conditions are fulfilled completely by specifying that

$$\delta_L(z) = (1+z)^n \qquad \text{and} \qquad \delta_H(z) = (1-z)^n. \tag{10}$$

On putting the specification of (10) into (7) and (8), we find that

$$\begin{aligned}\psi_L(z) &= \frac{(1+z)^n(1+z^{-1})^n}{(1+z)^n(1+z^{-1})^n + \lambda(1-z)^n(1-z^{-1})^n} \\ &= \frac{1}{1+\lambda\left(i\dfrac{1-z}{1+z}\right)^{2n}}\end{aligned} \tag{11}$$

and that

$$\begin{aligned}\psi_H(z) &= \frac{\lambda(1-z)^n(1-z^{-1})^n}{(1+z)^n(1+z^{-1})^n + \lambda(1-z)^n(1-z^{-1})^n} \\ &= \frac{1}{1+\dfrac{1}{\lambda}\left(i\dfrac{1+z}{1-z}\right)^{2n}}.\end{aligned} \tag{12}$$

These will be recognised as instances of the Butterworth digital filter which is familiar in electrical engineering—see, for example, Roberts and Mullis [15].

Since $\delta_L(z)$ and $\delta_H(z)$ are now completely specified, it follows that $\gamma(z)$ can be determined via the Cramér–Wold factorisation of the polynomial of (8). However, the virtue of the specification of (10), which places the zeros of the filters at $z = \pm 1$, is that it enables us to derive analytic expressions for the roots of the equation $\gamma(z)\gamma(z^{-1}) = 0$ which are the poles of the filters.

These roots come in reciprocal pairs; and, once they are available, they may be assigned unequivocally to the factors $\gamma(z)$ and $\gamma(z^{-1})$. Those roots which lie outside the unit circle belong to $\gamma(z)$ whilst their reciprocals, which lie inside the unit circle, belong to $\gamma(z^{-1})$.

Consider, therefore, the equation

$$(1+z)^n(1+z^{-1})^n + \lambda(1-z)^n(1-z^{-1})^n = 0. \tag{13}$$

which is equivalent to the equation

$$\lambda + \left(i\frac{1+z}{1-z}\right)^{2n} = 0. \tag{14}$$

Solving the latter for

$$s = i\frac{1+z}{1-z} \tag{15}$$

is a matter of finding the $2n$ roots of $-\lambda$. These are given by

$$s = \lambda^{1/(2n)} \exp\left\{\frac{i\pi j}{2n}\right\}, \quad \text{where} \quad j = 1, 3, 5, \ldots, 4n-1, \qquad \text{or} \quad j = 2k-1;\ k = 1, \ldots, 2n. \tag{16}$$

The roots correspond to a set of $2n$ points which are equally spaced around the circumference of a circle of radius λ. The radii which join the points to the centre are separated by angles of π/n; and the first of the radii makes an angle of $\pi/(2n)$ with the horizontal real axis.

The inverse of the function $s = s(z)$ is the function

$$z = \frac{s-i}{i+s} = \frac{(ss^* - 1) - i(s - s^*)}{(ss^* + 1) - i(s^* - s)}, \tag{17}$$

where the final expression, which has a real-valued denominator, comes from multiplying top and bottom of the second expression by $s^* - i = (i+s)^*$, where s^* denotes the conjugate of the complex number s. The elements of this formula are

$$\begin{aligned} ss^* &= \lambda^{1/n}, \\ s + s^* &= 2\lambda^{1/(2n)} \cos(\pi j/n) \quad \text{and} \\ i(s - s^*) &= 2\lambda^{1/(2n)} \sin(\pi j/n). \end{aligned} \tag{18}$$

The roots of $\gamma(z^{-1}) = 0$ are generated when $(j+1)/2 = k = 1, \ldots, n$. Those of $\gamma(z) = 0$ are generated when $k = n+1, \ldots, 2n$.

It is straightforward to determine the value of λ which will place the cut-off of the filter at a chosen point $\omega_c \in (0, \pi)$. Consider setting $z = \exp\{-i\omega\}$ in the formula of (11) the lowpass filter. This gives the following expression for the gain:

$$\begin{aligned} \psi_L(e^{-i\omega}) &= \frac{1}{1 + \lambda\left(i\dfrac{1 - e^{-i\omega}}{1 + e^{-i\omega}}\right)^{2n}} \\ &= \frac{1}{1 + \lambda\big\{\tan(\omega/2)\big\}^{2n}}. \end{aligned} \tag{19}$$

At the cut-off point, the gain must equal 1/2, whence solving the equation $\psi_L(\exp\{-i\omega_c\}) = 1/2$ gives $\lambda = \{1/\tan(\omega_c/2)\}^{2n}$.

Figure 1 shows the disposition in the complex plane of the poles and zeros of the factor $\delta_L(z^{-1})/\gamma_L(z^{-1})$ of the lowpass filter $\psi(z)_L$ for $n = 6$ when the cut-off point is at $\omega = \pi/2$ and at $\omega = \pi/8$. The poles are marked by crosses

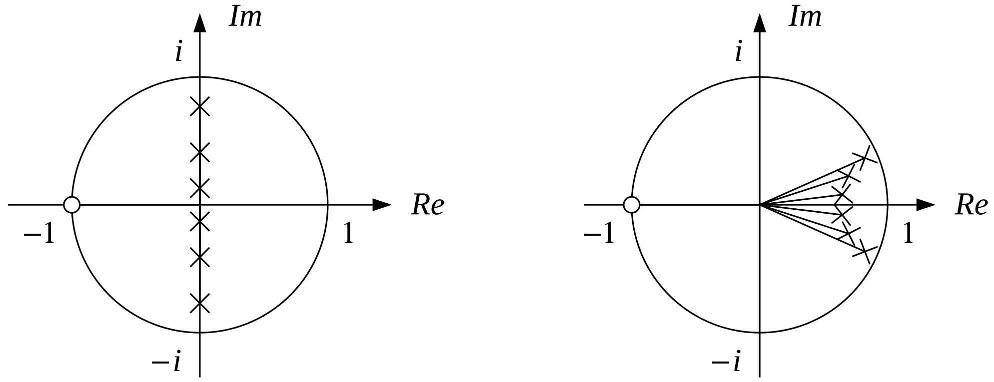

Figure 1. The pole–zero diagrams of the lowpass square-wave filters for $n = 6$ when the cut-off is at $\omega = \pi/2$ (left) and at $\omega = \pi/8$ (right).

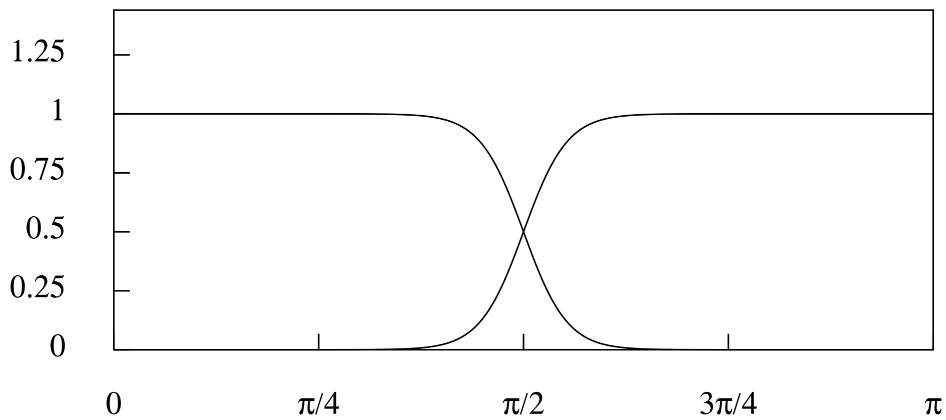

Figure 2. The frequency-responses of the prototype square-wave filters with $n = 6$ and with a cut-off at $\omega = \pi/2$.

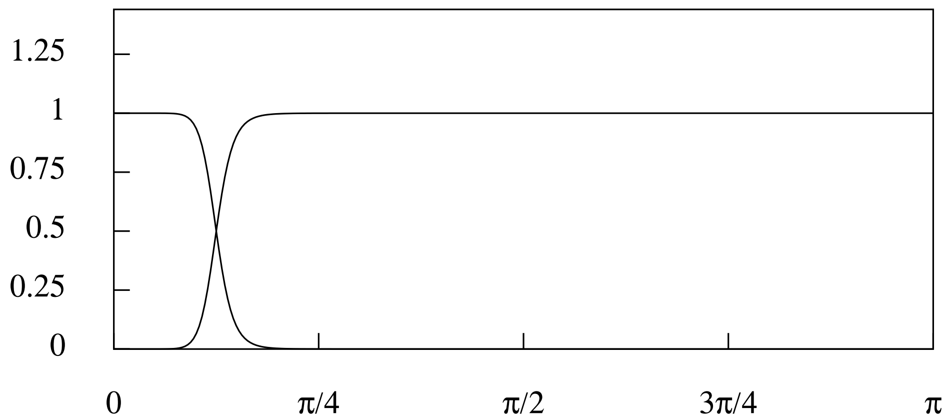

Figure 3. The frequency-responses of the square-wave filters with $n = 6$ and with a cut-off at $\omega = \pi/8$.

and the zeros by circles. Figures 2 and 3 show the gain of the filter for these two cases.

The rate of the transition from the passband to the stopband of a filter can be increased by increasing the value of n, which entails increasing the number of poles and the zeros. This is liable to bring some of the poles closer to the perimeter of the unit circle. It is also apparent from Figure 1 that, when the cut-off frequency is shifted away from the mid point $\pi/2$, a similar effect ensues whereby the rate of of transition is increased and the poles are brought closer to the perimeter.

When the poles are close to the perimeter of the unit circle, there can be problems of stability in implementing the filter. These can lead to the propagation of numerical rounding errors and to the prolongation of the transient effects of ill-chosen start-up conditions. Problems of numerical instability can be handled, within limits, by increasing the numerical precision with which the filter coefficients and the processed series are represented. The problems associated with the start-up conditions can be largely overcome by adopting the techniques which are described in the following section.

3. Implementing the Filters

The classical signal-extraction filters are intended to be applied to lengthy data sets which have been generated by stationary processes. The task of adapting them to limited samples from nonstationary processes often causes difficulties and perplexity. The problems arise from not knowing how to supply the initial conditions with which to start a recursive filtering process. By choosing inappropriate starting values for the forwards or the backwards pass, one can generate a so-called transient effect which is liable, in fact, to affect all of the processed values.

One common approach to the problem of the start-up conditions relies upon the ability to extend the sample by forecasting and backcasting. The additional extra-sample values can be used in a run-up to the filtering process wherein the filter is stabilised by providing it with a plausible history, if it is working in the direction of time, of with a plausible future, if it is working in reversed time. Sometimes, very lengthy extrapolations are called for (see Burman [2], for example).

An alternative approach to the start-up problem is to estimate the requisite initial conditions. Some of the methods which follow this approach have been devised within the context of the Kalman filter and the associated smoothing algorithms—see Ansley and Kohn [1] and de Jong [4], for example.

The approach which we shall adopt in this paper avoids the start-up problem by applying the filter, in the first instance, to a version of the data sequence which has been reduced to stationarity by repeated differencing. In this case, neither the differenced version of the trend nor the differenced version of the residual sequence require any start-up values other than the zeros which are the unconditional expectations of their elements.

We can proceed to find an estimate of the residual sequence by cumulating its differenced version. Here we can profit from some carefully estimated start-up values to set the process of cumulation in motion; but, in fact, we can avoid the finding such values explicitly. Moreover, these start-up values should tend to their zero-valued expectations as size of the sample increases; and, therefore, it would be acceptable, in a case of a lengthy data series, to set their values to zero.

Once the residual sequence has been estimated, the estimates of the trend sequence, which is its compliment within the data sequence, can be found by subtraction.

To clarify these matters, let us begin by considering a specific model for which the square-wave filter would represent the optimal device for extracting the trend, given a sample of infinite length. The model is represented by the equation

$$\begin{aligned} y(t) &= \xi(t) + \eta(t) \\ &= \frac{(1+L)^n}{(1-L)^d}\nu(t) + (1-L)^{n-d}\varepsilon(t), \end{aligned} \tag{20}$$

where $\xi(t)$ is the stochastic trend and $\eta(t)$ is the residual sequence. These are driven, respectively, by $\nu(t)$ and $\varepsilon(t)$ which are statistically independent sequences generated by normal white-noise processes with $V\{\nu(t)\} = \sigma_\nu^2$ and $V\{\varepsilon(t)\} = \sigma_\varepsilon^2$. Equation (20) can be rewritten as

$$\begin{aligned} (1-L)^d y(t) &= (1+L)^n\nu(t) + (1-L)^n\varepsilon(t), \\ &= \zeta(t) + \kappa(t) \\ &= g(t), \end{aligned} \tag{21}$$

where

$$\begin{aligned} &\zeta(t) = (1-L)^d\xi(t) = (1+L)^n\nu(t), \\ &\kappa(t) = (1-L)^d\eta(t) = (1-L)^n\varepsilon(t) \quad \text{and} \\ &g(t) = (1-L)^d y(t) = \zeta(t) + \kappa(t) \end{aligned} \tag{22}$$

are stationary moving-average processes which are noninvertible.

The Wiener–Kolmogorov theory of statistical signal extraction, as expounded by Whittle [17] for example, indicates that, when it is applied to the stationary series $g(t)$, the lowpass filter $\psi_L(z)$ of equation (11) will generate the minimum mean-square-error estimate of the sequence $\zeta(t)$ provided that the smoothing parameter has the value of $\lambda = \sigma_\varepsilon^2/\sigma_\nu^2$. More recently, Bell [2] has established that the Wiener–Kolmogorov theory applies equally to nonstationary processes. Thus, if it were applied to the nonstationary series $y(t)$, the lowpass filter would generate the minimum mean-square-error estimate of the nonstationary sequence $\xi(t)$.

The recursive filtering of nonstationary sequences is beset by two problems. On the one hand, there is the above-mentioned difficulty posed by the initial conditions. On the other hand, there is the danger that the unbounded nature of the sequence and the disparity of the values within it will lead to problems of numerical representation. Therefore, in pursuit of an alternative approach, we may consider the equation

$$\begin{aligned} \xi(t) &= y(t) - \eta(t) \\ &= y(t) - \frac{\kappa(t)}{(1-L)^d} \end{aligned} \tag{23}$$

which can be derived in reference to (20) and (22). Here $\kappa(t)$ is a stationary sequence which can be estimated by applying the highpass filter $\psi_H(L)$ of (12) to $g(t)$ which is the differenced version of the data sequence. Thereafter, an estimate of the stationary sequence $\eta(t)$ can be obtained by a d-fold process of cumulation which uses d zeros for its starting values—the zeros being the expected values of d consecutive elements of $\eta(t)$. Finally, the estimate $\xi(t)$ can be obtained by a simple subtraction.

Now imagine that, instead of a lengthy sequence of observations which can be treated as if it were infinite, there are only T observations of the process $y(t)$ of equation (20) which run from $t = 0$ to $t = T - 1$. These are gathered in a vector

$$\begin{aligned} y &= \xi + \eta \\ &= x + h, \end{aligned} \tag{24}$$

where ξ is the trend vector and η is the residual vector which is generated by a stationary process with

$$E(\eta) = 0 \qquad \text{and} \qquad D(\eta) = \sigma_\varepsilon^2 \Sigma. \tag{25}$$

The estimates of these vectors are denoted by x and h respectively.

To find the finite-sample the counterpart of equation (21), we need to represent the d-th difference operator $(1 - L)^d = 1 + \delta_1 L + \cdots + \delta_p L^d$ in the form of a matrix. The matrix which finds the d-th differences $g_d, \ldots, g_{T-1}$ of the data points $y_0, y_1, y_2, \ldots, y_{T-1}$ take the the form of

$$Q' = \begin{bmatrix} \delta_d & \ldots & \delta_1 & 1 & \ldots & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \ldots & \delta_d & \delta_{d-1} & \ldots & 1 & 0 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & \delta_d & \ldots & \delta_1 & 1 & \ldots & 0 & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & \delta_d & \delta_{d-1} & \ldots & 1 & 0 \\ 0 & \ldots & 0 & 0 & \ldots & 0 & \delta_d & \ldots & \delta_1 & 1 \end{bmatrix}. \tag{26}$$

Premultiplying equation (24) by this matrix gives

$$\begin{aligned} g = Q'y &= \zeta + \kappa \\ &= z + k, \end{aligned} \tag{27}$$

where $\zeta = Q'\xi$ and $\kappa = Q'\eta$ and were $z = Q'x$ and $k = Q'h$ are the corresponding estimates. The first and second moments of the vector ζ may be denoted by

$$E(\zeta) = 0 \qquad \text{and} \qquad D(\zeta) = \sigma_\nu^2 \Omega_L, \tag{28}$$

and those of κ by

$$\begin{aligned} E(\kappa) = 0 \qquad \text{and} \qquad D(\kappa) &= Q'D(\eta)Q \\ &= \sigma_\varepsilon^2 Q'\Sigma Q = \sigma_\varepsilon^2 \Omega_H, \end{aligned} \tag{29}$$

where both Ω_L and Ω_H are symmetric Toeplitz matrices with $2n + 1$ nonzero diagonal bands. The generating functions for the coefficients of these matrices are, respectively, $\delta_L(z)\delta_L(z^{-1})$ and $\delta_H(z)\delta_H(z^{-1})$, where $\delta_L(z)$ and $\delta_H(z)$ are the polynomials defined in (10). The generating function for the coefficients of Σ is $\{(1 - z)(1 - z^{-1})\}^{n-d}$.

The optimal predictor z of the vector $\zeta = Q'\xi$ is given by the following conditional expectation:

$$\begin{aligned} E(\zeta|g) &= E(\zeta) + C(\zeta, g)D^{-1}(g)\{g - E(g)\} \\ &= \Omega_L(\Omega_L + \lambda\Omega_H)^{-1}g = z, \end{aligned} \tag{30}$$

where $\lambda = \sigma_\varepsilon^2/\sigma_\nu^2$. The optimal predictor k of $\kappa = Q'\eta$ is given, likewise, by

$$\begin{aligned} E(\kappa|g) &= E(\kappa) + C(\kappa, g)D^{-1}(g)\{g - E(g)\} \\ &= \lambda\Omega_H(\Omega_L + \lambda\Omega_H)^{-1}g = k. \end{aligned} \tag{31}$$

It may be confirmed that $z + k = g$.

The estimates are calculated, first, by solving the equation

$$(\Omega_L + \lambda\Omega_H)b = g \tag{32}$$

for the value of b and, thereafter, by finding

$$z = \Omega_L b \qquad \text{and} \qquad k = \lambda\Omega_H b. \tag{33}$$

The solution of equation (32) is found via a Cholesky factorisation which sets $\Omega_L + \lambda\Omega_H = GG'$, where G is a lower-triangular matrix. The system $GG'b = g$

may be cast in the form of $Gp = g$ and solved for p. Then $G'b = p$ can be solved for b.

Observe that the generating function for the matrix GG' is the polynomial $\gamma(z)\gamma(z^{-1})$ defined in (8). Moreover, the solution via the Cholesky factorisation is a finite sample analogue of a process of bidirectional filtering which finds the sequence $b(t) = \{\gamma(F)\gamma(L)\}^{-1}g(t)$ via the recursive processes $\gamma(L)p(t) = g(t)$ and $\gamma(F)b(t) = p(t)$. The equations of (33) correspond, respectively, to the filtering operations $z(t) = \delta_L(F)\delta_L(L)b(t)$ and $k(t) = \delta_H(F)\delta_H(L)b(t)$, each of which can be realised in a single pass.

Our object is to recover from z or from k an estimate x of the trend vector ξ. This would be conceived, ordinarily, as a matter of integrating the vector z d times via a simple recursion which depends upon d initial conditions. However we can circumvent the problem of finding the initial conditions by seeking the solution to the following problem:

$$\text{(34)} \qquad \text{Minimise} \quad (y-x)'\Sigma^{-1}(y-x) \quad \text{subject to} \quad Q'x = z.$$

The problem is addressed by evaluating the Lagrangean function

$$\text{(35)} \qquad L(x,\mu) = (y-x)'\Sigma^{-1}(y-x) + 2\mu'(Q'x - z).$$

By differentiating the function with respect to x and setting the result to zero, we obtain the condition

$$\text{(36)} \qquad \Sigma^{-1}(y-x) - Q\mu = 0.$$

Premultiplying by $Q'\Sigma$ gives

$$\text{(37)} \qquad Q'(y-x) = Q'\Sigma Q\mu.$$

But, from (32) and (33), it follows that

$$\text{(38)} \qquad \begin{aligned} Q'(y-x) &= g - z \\ &= \lambda\Omega_H b = \lambda Q'\Sigma Q b, \end{aligned}$$

whence, from (37), we get

$$\text{(39)} \qquad \begin{aligned} \mu &= (Q'\Sigma Q)^{-1}Q'(y-x) \\ &= \lambda b. \end{aligned}$$

Putting the final expression for μ into (36) gives

$$\text{(40)} \qquad x = y - \lambda\Sigma Q b.$$

This is our solution to the problem of estimating the trend vector ξ.

Observe that

$$
\begin{aligned}
h &= \lambda\Sigma Q b \\
&= \Sigma Q(Q'\Sigma Q)^{-1}k
\end{aligned} \tag{41}
$$

is the estimate of the vector η. Moreover it is subject to an identity $Q'h = k$ which corresponds to the identity $Q'\eta = \kappa$ which defines κ. In fact, the difference between the mapping from k to h in this equation and the simple d-fold accumulation of the elements of k which is suggested by equation (23) disappears as the size of the sample increases.

Notice also that, if y is a vector of T values of a polynomial of degree $d-1$ taken at equally spaced values, then $g = Q'y = 0$ and $b = 0$, whence $x = y$. Thus a polynomial time trend of degree less than d is unaffected by the filter. This is indeed the appropriate outcome; and it would not be forthcoming if the startup problem were handled in another way.

It is notable that there is a criterion function which will enable us to derive the equation of the trend estimation filter in a single step. The function is

$$
L(x) = (y-x)'\Sigma^{-1}(y-x) + \lambda x'Q\Omega_L^{-1}Q'x, \tag{42}
$$

wherein $\lambda = \sigma_\varepsilon^2/\sigma_\nu^2$ as before. This is minimised by the value specified in (40). The criterion function becomes intelligible when we allude to the assumptions that $y \sim N(\xi, \sigma_\varepsilon^2\Sigma)$ and that $Q'\xi = \zeta \sim N(0, \sigma_\nu^2\Omega_L)$; for then it plainly resembles a sum of two independent chi-square variates scaled by a factor of σ_ε^2.

4. An Application of the Rational Filter

We shall illustrate the uses of the rational filter by applying it to a series of 66 figures which constitute the quarterly unemployment statistics for Switzerland from the first quarter of 1980 through to the second quarter of 1996. The graph of this series is given in Figure 4. Our objective is to discover the pattern of the seasonal fluctuations which surround the longer-term trend. A casual inspection of the graph would suggest that the seasonal motions have been in abeyance in the period of rapidly increasing unemployment in the third segment of the series, only to be resumed when unemployment is stabilised at a higher level at the end of the series.

The angular velocity of the seasonal fluctuation is $\pi/2$ radians, or 90 degrees, per period; and our objective of removing the trend would be fulfilled by eliminating every component of a lesser frequency. In fact, we shall choose a nominal cut-off point for the filter of $3\pi/8$ radians, or 67.5 degrees. This places the transition between the pass band and the stop band in an area which corresponds to a dead space in the periodogram of the data where there are no components of any significant power. The existence of this dead space allows

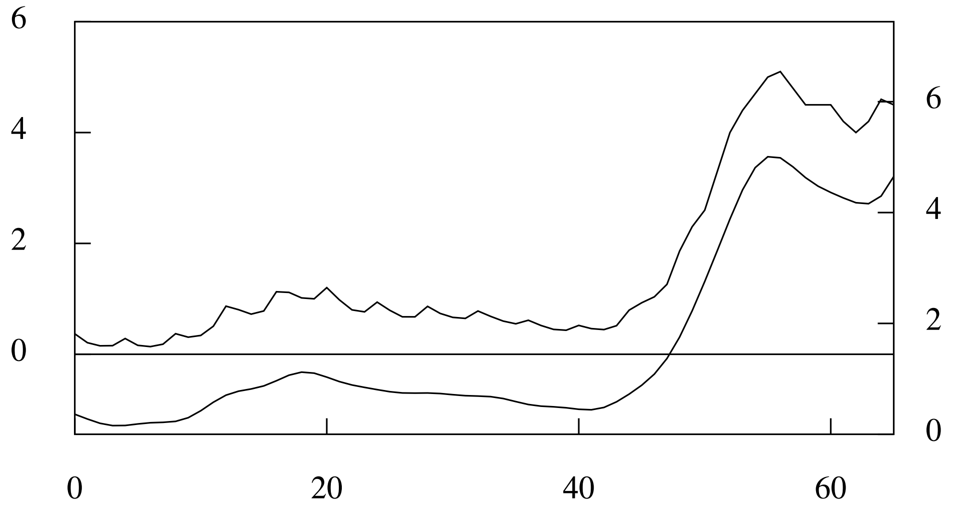

Figure 4. The quarterly figures on Swiss unemployment from 1980.1 to 1996.2 (upper panel) together with their trend obtained by smoothing the series using a lowpass filter of order $n = 8$.

us to use a filter of relatively low orders which has a more gradual transition than would be tolerated in more exacting circumstances. The effects of the choices of the cut-off point and the filter orders can be seen in Figure 5.

Figure 6 shows the residuals of the series after the trend has been extracted using the lowpass filter. What is remarkable about this series is its regularity. The amplitudes of the seasonal fluctuations are clearly related to the level of unemployment. Thus, in times of high employment, there appears to be a widespread hoarding of labour which would be subject, at other times, to seasonal unemployment. This is a feature which one would not have detected by inspecting the original data series. It also transpires, from Figure 6, that, far from being in abeyance during the period of rapidly increasing unemployment, the seasonal fluctuations were present and were of a steadily increasing amplitude.

The regularity of the residual series is reflected in its periodogram which is represented in Figure 7. Here, the complete absence of any components of a frequency below the cut-off point is a powerful testimony to the efficacy of the rational filter. The tall spike centred at 90 degrees, or $\pi/2$ radians, represents the power of the seasonal fluctuations.

The effects of a parallel analysis of the unemployment figures which has used the Hodrick–Prescott filter are represented in Figures 8 to 10. The filter fails to remove from the residual sequence some of the motions which ought to be attributed to the trend. The consequences is that the regularity of the seasonal effect in not apparent in the residual sequence and the false impression is strengthened that the effect is largely in abeyance during the period of the

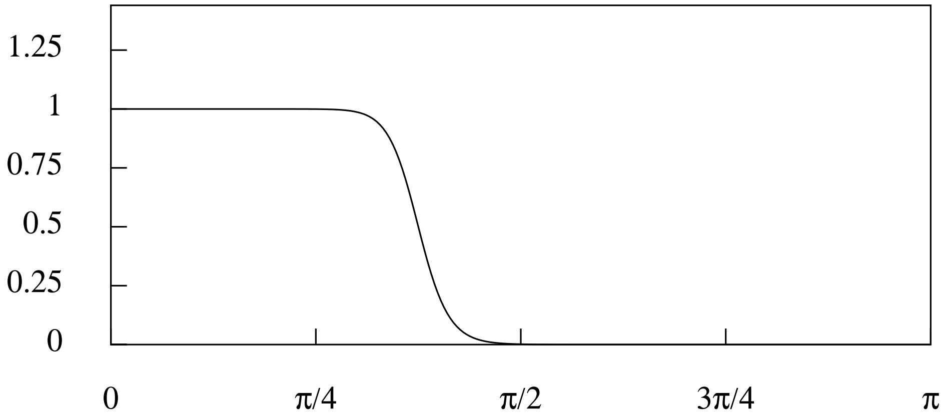

Figure 5. The gain of the lowpass filter of order $n = 8$ with the nominal cut-off point at $\omega = 3\pi/8$.

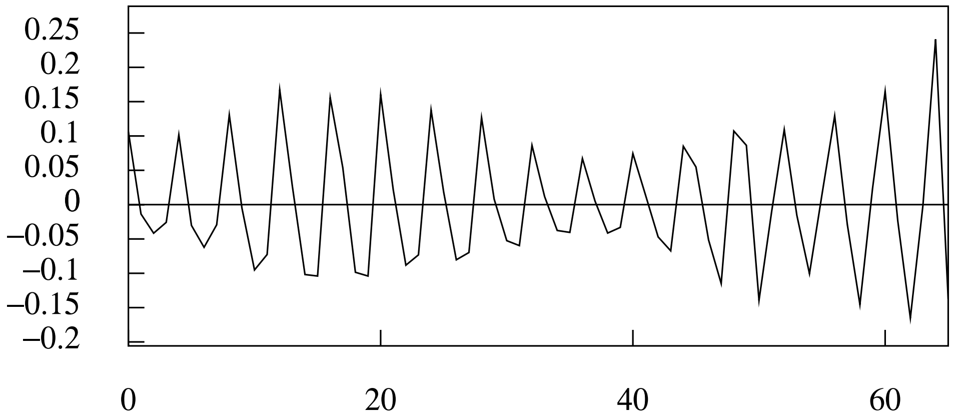

Figure 6. The residual sequence from detrending the Swiss unemployment figures using the rational filter.

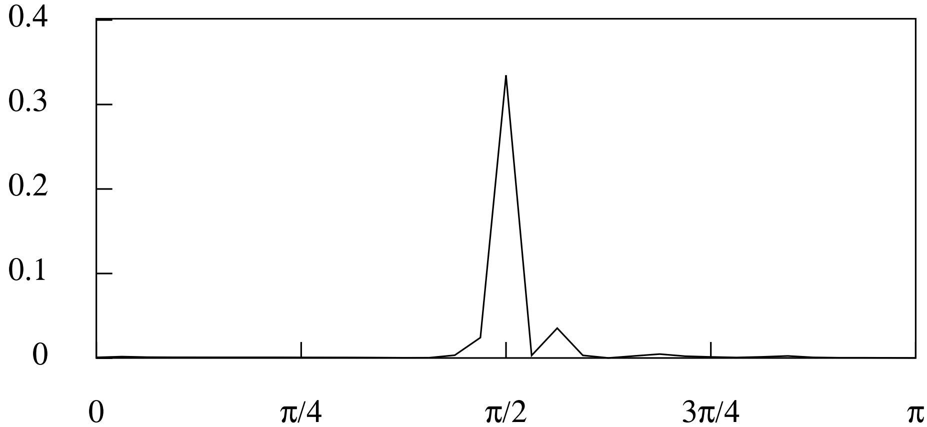

Figure 7. The periodogram of the residual sequence obtained by detrending the Swiss unemployment figures using the rational filter.

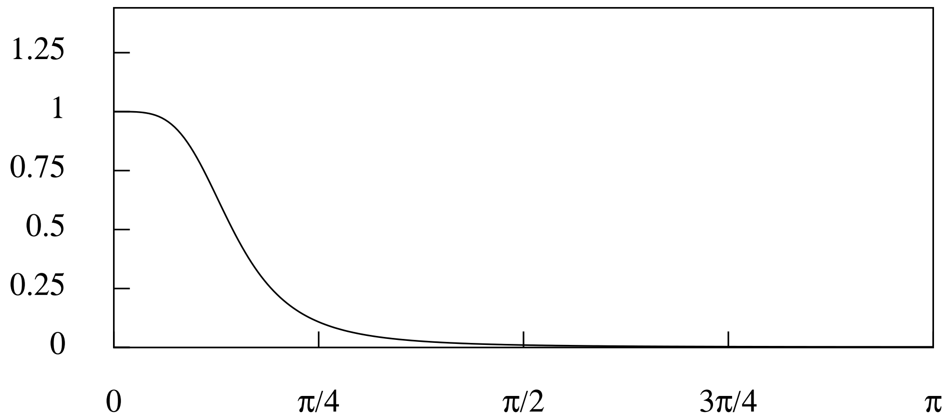

Figure 8. The gain of the Hodrick–Prescott lowpass filter with a smoothing parameter of 24.

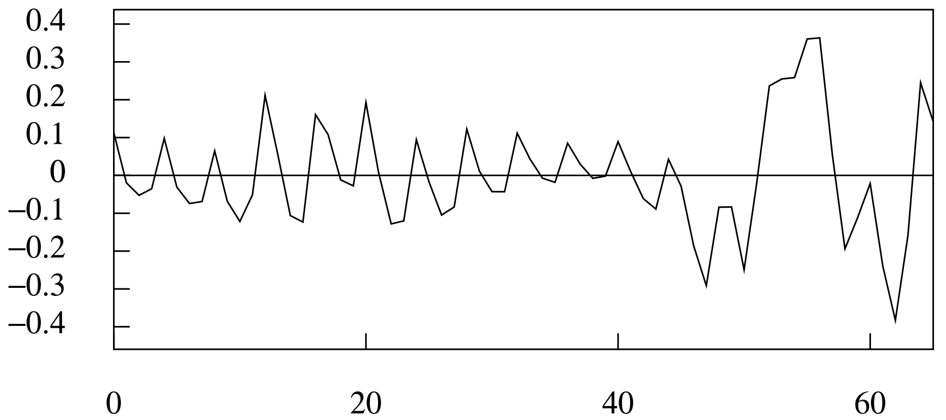

Figure 9. The residual sequence from detrending the Swiss unemployment figures using the Hodrick–Prescott filter.

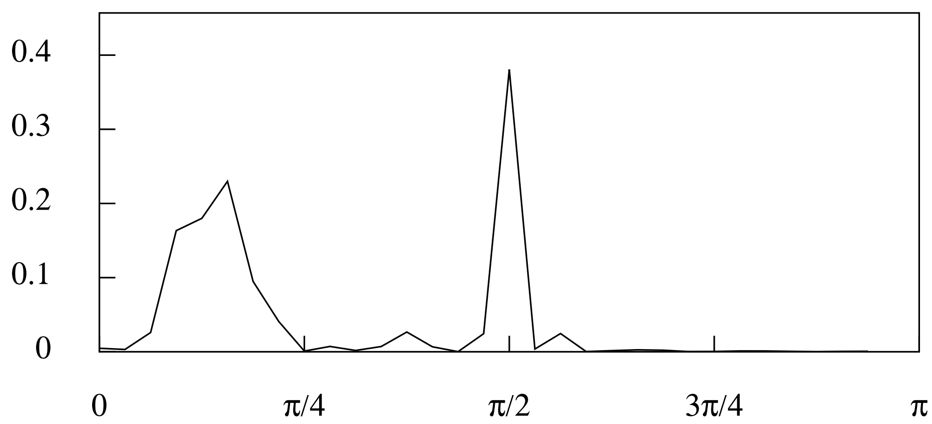

Figure 10. The periodogram of the residual sequence obtained by detrending the Swiss unemployment figures using the Hodrick–Prescott filter.

rapid increase in unemployment. The fault of the filter is evident in Figure 10 which shows that it has allowed some powerful low-frequency components to pass through into the residual series.

This example shows that is is sometimes possible to recover detailed structural information which is buried within aggregate time series; and it emphasises the need, in analysing economic time series, to use frequency-selective filters with well-defined cut-off points.

References

[1] Ansley, C.F., and R. Kohn, (1985), Estimation, Filtering and Smoothing in State Space Models with Incompletely Specified Initial Conditions, *The Annals of Statistics,* **13,** 1286–1316.

[2] Bell, W., (1984), Signal Extraction for Nonstationary Time Series, *The Annals of Statistics,* **12,** 646–664.

[3] Burman, J.P., (1980), Seasonal Adjustment by Signal Extraction, *Journal of the Royal Statistical Society, Series A,* **143,** 321–337.

[4] Cogley, T., and J.M. Nason, (1995), Effects of the Hodrick–Prescott Filter on Trend and Difference Stationary Time Series, Implications for Business Cycle Research, *Journal of Economic Dynamics and Control,* **19,** 253–278.

[5] De Jong, P., (1991), The Diffuse Kalman Filter, *The Annals of Statistics,* **19,** 1073–1083.

[6] Hartley, J.E., K.D. Hoover and K.D. Salyer, (eds.) (1998), *Real Business Cycles: A Reader,* Routledge, London.

[7] Harvey, A.C., and A. Jaeger, (1993), Detrending, Stylised Facts and the Business Cycle, *Journal of Applied Econometrics,* **8,** 231–247.

[8] Harvey, A.C., and P.H. Todd, (1983), Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study, *Journal of Business and Economic Forecasting,* **1,** 299–307.

[9] Hillmer, S.C., and G.C. Tiao, (1982), An ARIMA-Model-Based Approach to Seasonal Adjustment, *Journal of the American Statistical Association,* **77,** 63–70.

[10] King, R.G., and S.G. Rebelo, (1993), Low Frequency Filtering and Real Business Cycles, *Journal of Economic Dynamics and Control,* **17,** 207–231.

[11] Kolmogorov, A.N., (1941), Interpolation and Extrapolation, *Bulletin de l'academie des sciences de U.S.S.R., Ser. Math.,* **5,** 3–14.

[12] Koopman, S.J., A.C. Harvey, J.A. Doornick and N. Shephard, (1995), *STAMP 5.0: Structural Time Series Analyser Modeller and Predictor—The Manual,* Chapman and Hall, London.

[13] Kydland, F.E., and C. Prescott, (1990), Business Cycles: Real Facts and a Monetary Myth, *Federal Reserve Bank of Minneapolis Quarterly Review,* **14,** 3–18.

[14] Reinsch, C.H., (1976), Smoothing by Spline Functions, *Numerische Mathematik,* **10,** 177–183.

[15] Roberts, R.A., and C.T. Mullis, (1987), *Digital Signal Processing,* Addison Wesley, Reading, Massachusetts.

[16] Wiener, N., (1950), *Extrapolation, Interpolation and Smoothing of Stationary Time Series,* MIT Technology Press and John Wiley and Sons, New York.

[17] Whittle, P., (1983), *Prediction and Regulation by Linear Least-Square Methods, Second Revised Edition,* Basil Blackwell, Oxford.