WIENER–KOLMOGOROV FILTERING, FREQUENCY-SELECTIVE FILTERING AND POLYNOMIAL REGRESSION

by D.S.G. POLLOCK

Queen Mary, University of London

Email: stephen_pollock@sigmapi.u-net.com

Adaptations of the classical Wiener–Kolmogorov filters are described that enable them to be applied to short nonstationary sequences. Alternative filtering methods that operate in the time domain and the frequency domain are described. The frequency-domain methods have the advantage of allowing components of the data to be separated along sharp dividing lines in the frequency domain, without incurring any leakage. The paper contains a novel treatment of the start-up problem that affects the filtering of trended data sequences.

Date: February, 2006

Key words: Linear Filters, Spectral Analysis, Signal Extraction.

JEL Classification: C22

1. INTRODUCTION

The classical theory of statistical signal extraction presupposes lengthy data sequences, which are assumed, in theory, to be doubly infinite or semi-infinite—see Whittle (1983), for example. In many practical cases, and in most econometric applications, the available data are, to the contrary, both strongly trended and of a limited duration.

This paper is concerned with the theory of finite-sample signal extraction; and it shows how the classical Wiener–Kolmogorov theory of signal extraction can be adapted to cater to short sequences generated by processes that may be nonstationary.

Alternative methods of processing finite samples, which work in the frequency domain, are also described; and their relation to the time-domain implementations of the Wiener–Kolmogorov methods are demonstrated. The frequency-domain methods have the advantage that they are able to achieve clear separations of components that reside in adjacent frequency bands in a way that the time-domain methods cannot.

2. WIENER–KOLMOGOROV FILTERING OF SHORT STATIONARY SEQUENCES

In the classical theory, it is assumed that there is a doubly-infinite sequence of observations, denoted, in this paper, by $y(t) = \{y_t; t = 0, \pm 1, \pm 2, \ldots\}$. Here, we shall assume that the observations run from $t = 0$ to $t = T - 1$. These are gathered in the vector $y = [y_0, y_1, \ldots, y_{T-1}]'$, which is decomposed as

$$y = \xi + \eta, \tag{1}$$

where ξ is the signal component and η is the noise component. It may be assumed that the latter are from independent zero-mean Gaussian processes that are completely characterised by their first and second moments. Then,

$$\begin{aligned} E(\xi) &= 0, \qquad D(\xi) = \Omega_\xi, \\ E(\eta) &= 0, \qquad D(\eta) = \Omega_\eta, \\ \text{and} \quad C(\xi, \eta) &= 0. \end{aligned} \tag{2}$$

A consequence of the independence of ξ and η is that $D(y) = \Omega = \Omega_\xi + \Omega_\eta$.

The autocovariance or dispersion matrices, which have a Toeplitz structure, may be obtained by replacing the argument z within the relevant autocovariance generating functions by the matrix

$$L_T = [e_1, \ldots, e_{T-1}, 0], \tag{3}$$

which is derived from the identity matrix $I_T = [e_0, e_1, \ldots, e_{T-1}]$ by deleting the leading column and appending a column of zeros to the end of the array. Using L_T in place of z in the autocovariance generating function $\gamma(z)$ of the data process gives

$$D(y) = \Omega = \gamma_0 I_T + \sum_{\tau=1}^{T-1} \gamma_\tau (L_T^\tau + F_T^\tau), \tag{4}$$

where $F_T = L_T'$ is in place of z^{-1}. Since L_T and F_T are nilpotent of degree T, such that $L_T^q, F_T^q = 0$ when $q \geq T$, the index of summation has an upper limit of $T - 1$.

The optimal predictors of the signal and the noise components are

$$\begin{aligned} E(\xi|y) &= E(\xi) + C(\xi, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y = Z_\xi y = x, \end{aligned} \tag{5}$$

$$\begin{aligned} E(\eta|y) &= E(\eta) + C(\eta, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y = Z_\eta y = h, \end{aligned} \tag{6}$$

which are their minimum-mean-square-error estimates.

The corresponding error dispersion matrices, from which confidence intervals for the estimated components may be derived, are

$$\begin{aligned} D(\xi|y) &= D(\xi) - C(\xi, y)D^{-1}(y)C(y, \xi) \\ &= \Omega_\xi - \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\xi, \end{aligned} \tag{7}$$

$$\begin{aligned} D(\eta|y) &= D(\eta) - C(\eta, y)D^{-1}(y)C(y, \eta), \\ &= \Omega_\eta - \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\eta. \end{aligned} \tag{8}$$

These formulae contain $D\{E(\xi|y)\} = C(\xi, y)D^{-1}(y)C(y, \xi)$ and $D\{E(\eta|y)\} = C(\eta, y)D^{-1}(y)C(y, \eta)$, which give the variability of the estimated components relative to their zero-valued unconditional expectations. The results follow from

the ordinary algebra of conditional expectations, of which an account has been given by Pollock (1999). (Equations (5) and (6) are the basis of the analysis in Pollock 2000.)

It can be seen, from (5) and (6) that $y = x + h$. Therefore, only one of the components needs be calculated. The estimate of the other component may be obtained by subtracting the calculated estimate from y. Also, the matrix inversion lemma indicates that

$$\begin{aligned}(\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1} &= \Omega_\eta - \Omega_\eta(\Omega_\eta + \Omega_\xi)^{-1}\Omega_\eta \\ &= \Omega_\xi - \Omega_\xi(\Omega_\eta + \Omega_\xi)^{-1}\Omega_\xi.\end{aligned} \tag{9}$$

Therefore, (7) and (8) represent the same quantity, which is to be expected in view of the adding up.

The estimating equations can be obtained via the criterion

$$\text{Minimise} \quad S(\xi, \eta) = \xi'\Omega_\xi^{-1}\xi + \eta'\Omega_\eta^{-1}\eta \quad \text{subject to} \quad \xi + \eta = y. \tag{10}$$

Since $S(\xi, \eta)$ is the exponent of the normal joint density function $N(\xi, \eta)$, the estimates of (5) and (6) may be described, alternatively, as the minimum chi-square estimates or as the maximum-likelihood estimates.

Setting $\eta = y - \xi$ gives the function

$$S(\xi) = \xi'\Omega_\xi^{-1}\xi + (y - \xi)'\Omega_\eta^{-1}(y - \xi). \tag{11}$$

The minimising value is

$$\begin{aligned}x &= (\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1}\Omega_\eta^{-1}y \\ &= y - \Omega_\eta(\Omega_\eta + \Omega_\xi)^{-1}y = y - h,\end{aligned} \tag{12}$$

where the second equality depends upon the first of the identities of (9).

A simple procedure for calculating the estimates x and h begins by solving the equation

$$(\Omega_\xi + \Omega_\eta)b = y \tag{13}$$

for the value of b. Thereafter, one can generate

$$x = \Omega_\xi b \qquad \text{and} \qquad h = \Omega_\eta b. \tag{14}$$

If Ω_ξ and Ω_η correspond to the dispersion matrices of moving-average processes, then the solution to equation (13) may be found via a Cholesky factorisation that sets $\Omega_\xi + \Omega_\eta = GG'$, where G is a lower-triangular matrix with a limited number of nonzero bands. The system $GG'b = y$ may be cast in the form of $Gp = y$ and solved for p. Then, $G'b = p$ can be solved for b. The procedure has been described by Pollock (2000).

3. FILTERING VIA FOURIER METHODS

A circular stochastic process of order T is one in which the autocovariance matrix of a random vector $x = [x_0, x_1, \ldots, x_{T-1}]'$, generated by the process,

is unchanged when the elements of x are subjected to a cyclical permutation. Such a process is the finite equivalent of a stationary stochastic process.

The circular permutation of the elements of the sequence is effected by the matrix operator

$$K_T = [e_1, \ldots, e_{T-1}, e_0], \tag{15}$$

which is formed from the identity matrix I_T by moving the leading column to the back of the array. The powers of K_T are T-periodic such that $K_T^{q+T} = K_T^q$. Moreover, any circular matrix of order T can be expressed as a linear combination of the basis matrices $I_T, K_T, \ldots, K_T^{T-1}$.

The matrix K_T has a spectral factorisation that entails the discrete Fourier transform and its inverse. Let

$$U = T^{-1/2}[W^{jt}; t, j = 0, \ldots, T-1] \tag{16}$$

denote the symmetric Fourier matrix, of which the generic element in the jth row and tth column is

$$\begin{aligned} W^{jt} &= \exp(-\mathrm{i}2\pi tj/T) = \cos(\omega_j t) - \mathrm{i}\sin(\omega_j t), \\ &\text{where} \quad \omega_j = 2\pi j/T. \end{aligned} \tag{17}$$

The matrix U is unitary, which is to say that it fulfils the condition

$$\bar{U}U = U\bar{U} = I_T, \tag{18}$$

where $\bar{U} = T^{-1/2}[W^{-jt}; t, j = 0, \ldots, T-1]$ denotes the conjugate matrix.

The circulant lag operator can be factorised as

$$K_T = \bar{U}DU = \bar{U}\bar{D}U, \tag{19}$$

where

$$D = \mathrm{diag}\{1, W, W^2, \ldots, W^{T-1}\} \tag{20}$$

is a diagonal matrix whose elements are the T roots of unity, which are found on the circumference of the unit circle in the complex plane. Observe also that D is T-periodic, such that $D^{q+T} = D^q$, and that $K^q = \bar{U}D^qU = \bar{U}\bar{D}^qU$ for any integer q. (An account of the algebra of circulant matrices has been provided by Pollock 2002. See, also, Gray 2002.)

The matrix of the circular autocovariances of the data is obtained by replacing the argument z in the autocovariance generating function $\gamma(z)$ by the matrix K_T:

$$\begin{aligned} D^\circ(y) = \Omega^\circ &= \gamma(K_T) \\ &= \gamma_0 I_T + \sum_{\tau=1}^{\infty} \gamma_\tau (K_T^\tau + K_T^{-\tau}) \\ &= \gamma_0^\circ I_T + \sum_{\tau=1}^{T-1} \gamma_\tau^\circ (K_T^\tau + K_T^{-\tau}). \end{aligned} \tag{21}$$

The circular autocovariances would be obtained by wrapping the sequence of ordinary autocovariances around a circle of circumference T and adding the overlying values. Thus

$$\gamma_\tau^\circ = \sum_{j=0}^{\infty} \gamma_{jT+\tau}, \quad \text{with} \quad \tau = 0, \ldots, T-1. \tag{22}$$

Given that $\lim(\tau \to \infty)\gamma_\tau = 0$, it follows that $\gamma_\tau^\circ \to \gamma_\tau$ as $T \to \infty$, which is to say that the circular autocovariances converge to the ordinary autocovariances as the circle expands.

The circulant autocovariance matrix is amenable to a spectral factorisation of the form

$$\Omega^\circ = \gamma(K_T) = \bar{U}\gamma(D)U, \tag{23}$$

wherein the jth element of the diagonal matrix $\gamma(D)$ is

$$\gamma(\exp\{\mathrm{i}\omega_j\}) = \gamma_0 + 2\sum_{\tau=1}^{\infty} \gamma_\tau \cos(\omega_j \tau). \tag{24}$$

This represents the cosine Fourier transform of the sequence of the ordinary autocovariances; and it corresponds to an ordinate (scaled by 2π) sampled at the point $\omega_j = 2\pi j/T$, which is a Fourier frequency, from the spectral density function of the linear (i.e. non-circular) stationary stochastic process.

The method of Wiener–Kolmogorov filtering can also be implemented using the circulant dispersion matrices that are given by

$$\begin{aligned} &\Omega_\xi^\circ = \bar{U}\gamma_\xi(D)U, \quad \Omega_\eta^\circ = \bar{U}\gamma_\eta(D)U \quad \text{and} \\ &\Omega^\circ = \Omega_\xi^\circ + \Omega_\eta^\circ = \bar{U}\{\gamma_\xi(D) + \gamma_\eta(D)\}U, \end{aligned} \tag{25}$$

wherein the diagonal matrices $\gamma_\xi(D)$ and $\gamma_\eta(D)$ contain the ordinates of the spectral density functions of the component processes. By replacing the dispersion matrices within (5) and (6) by their circulant counterparts, we derive the following formulae:

$$x = \bar{U}\gamma_\xi(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}Uy = P_\xi y, \tag{26}$$

$$h = \bar{U}\gamma_\eta(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}Uy = P_\eta y. \tag{27}$$

Similar replacements within the formulae (7) and (8) provide the expressions for the error dispersion matrices that are appropriate to the circular filters.

The filtering formulae may be implemented in the following way. First, a Fourier transform is applied to the data vector y to give Uy, which resides in the frequency domain. Then, the elements of the transformed vector are multiplied by those of the diagonal weighting matrices $J_\xi = \gamma_\xi(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}$ and $J_\eta = \gamma_\eta(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}$. Finally, the products are carried back into the time domain by the inverse Fourier transform, which is represented by the matrix $\bar{U}$. (An efficient implementation of a mixed-radix fast Fourier transform,

which is designed to cope with samples of arbitrary sizes, has been provided by Pollock (1999). The usual algorithms demand a sample size of $T = 2^n$.)

The frequency-domain realisations of the Wiener–Kolmogorov filters have sufficient flexibility to accommodate cases where the component processes $\xi(t)$ and $\eta(t)$ have band-limited spectra that are zero-valued beyond certain bounds. If the bands do not overlap, then it is possible to achieve a perfect decomposition of $y(t)$ into its components.

Let $\Omega_\xi^\circ = \bar{U}\Lambda_\xi U$, $\Omega_\eta^\circ = \bar{U}\Lambda_\eta U$ and $\Omega^\circ = \bar{U}(\Lambda_\xi + \Lambda_\eta)U$, where Λ_ξ and Λ_η contain the ordinates of the spectral density functions of $\xi(t)$ and $\eta(t)$, sampled at the Fourier frequencies. Then, if these spectra are disjoint, there will be $\Lambda_\xi\Lambda_\eta = 0$, and the dispersion matrices of the two processes will be singular. The matrix $\Omega_y^\circ = \Omega_\xi^\circ + \Omega_\eta^\circ$ will also be singular, unless the domains of the spectral density functions of the component processes partition the frequency range. Putting these details into (26) gives

$$x = \bar{U}\Lambda_\xi\{\Lambda_\xi + \Lambda_\eta\}^+ Uy = \bar{U}J_\xi Uy, \tag{28}$$

where $\{\Lambda_\xi + \Lambda_\eta\}^+$ denotes a generalised inverse. The corresponding error dispersion matrix is

$$\Omega_\xi^\circ - \Omega_\xi^\circ(\Omega_\xi^\circ + \Omega_\eta^\circ)^+\Omega_\xi^\circ = \bar{U}\Lambda_\xi U - \bar{U}\Lambda_\xi(\Lambda_\xi + \Lambda_\eta)^+\Lambda_\xi U. \tag{29}$$

But, if $\Lambda_\xi\Lambda_\eta = 0$, then $\Lambda_\xi(\Lambda_\xi + \Lambda_\eta)^+\Lambda_\xi = \Lambda_\xi$; and so the error dispersion is manifestly zero, which implies that $x = \xi$.

The significant differences between the time-domain and frequency-domain realisations of the Wiener–Kolmogorov filters occur at the ends of the sample. In the time-domain version, the coefficients of the filter vary as it moves through the sample, in a manner that prevents the filter from reaching beyond the bounds of the sample.

In the frequency-domain version, the implicit filter coefficients, which do not vary, are applied to the sample via a process of circular convolution. Thus, as it approaches the end of the sample, the filter takes an increasing proportion of its data from the beginning.

The implicit filter coefficients of the frequency-domain method are the circularly wrapped versions of the coefficients of a notional time-domain filter, defined in respect of $y(t) = \{y_t; t = 0, \pm1, \pm2, \ldots\}$, which may have an infinite impulse response. As the circle expands, the wrapped coefficients converge on the original coefficients in the same way as the circular autocovariances converge on the ordinary autocovariances.

Even when they form an infinite sequence, the coefficients the notional time-domain filter will have a limited dispersion around their central value. Thus, as the sample size increases, the proportion of the filtered sequence that suffers from the end-effects diminishes, and the central values from the time-domain filtering, realised according to the formulae (5) and (6), and the frequency-domain filtering, realised according to (26) and (27), will converge point on point. By extrapolating the sample by a short distance at either end, one should be able to place the end-effects of the frequency-domain filter out of range of the sample.

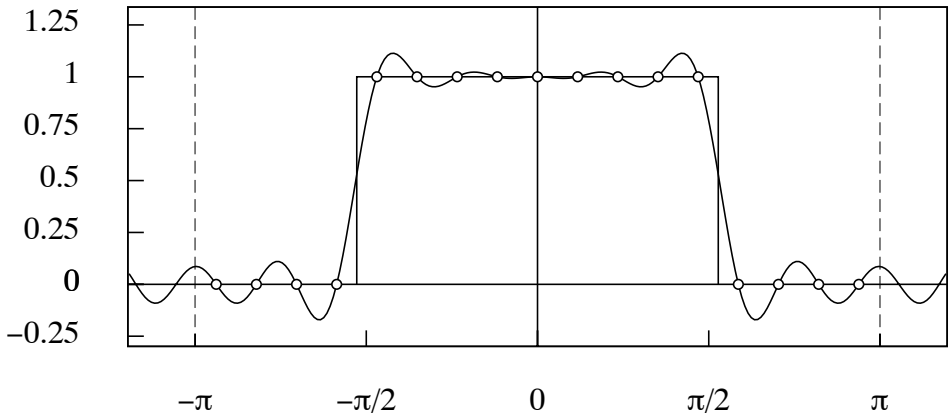

Figure 1. The frequency response of the 17-point wrapped filter defined over the interval $[-\pi, \pi)$. The values at the Fourier frequencies are marked by circles.

Example. The frequency response of a circular filter applied to a finite sample of T data points is a periodic function defined over a set of T adjacent Fourier frequencies $\omega_j = 2\pi j/T$, where j takes T consecutive integer values. It is convenient to set $j = 0, 1, \ldots, T-1$, in which case, the response is defined over the frequency interval $[0, 2\pi)$.

However, it may be required to define the response over the interval $[-\pi, \pi)$, which is used, conventionally, for representing the continuous response function of a filter applied to a doubly-infinite data sequence. In that case, $j = 0, \pm 1, \ldots, [T/2]$, where $[T/2]$ is the integral part of $T/2$.

The responses at the T Fourier frequencies can take any finite real values, subject only to the condition that $\lambda_j = \lambda_{T-j}$, or to the equivalent condition that $\lambda_j = \lambda_{-j}$. These restrictions will ensure that the corresponding filter coefficients form a real-valued symmetric sequence. The T coefficients of the circular filter may be obtained by applying the inverse discrete Fourier transform to the responses.

It is convenient, nevertheless, to specify the frequency response of a circular filter by sampling a pre-defined continuous response. Then, the coefficients of the circular filter are formally equivalent to those that would be found by wrapping the coefficients of an ordinary filter, pertaining to the continuous response, around a circle of circumference T and adding the coincident values.

A filter that has a band-limited frequency response has an infinite set of filter coefficients. In that case, it is not practical to find coefficients of the circular filter by wrapping the coefficients of the ordinary filter. Instead, they must be found by transforming a set of frequency-domain ordinates.

A leading example of a band-limited response is provided by an ideal low-pass frequency-selective filter that has a boxcar frequency response. Two versions of the filter may be defined, depending on whether the transitions between pass band and stop band occur at Fourier frequencies or between Fourier frequencies.

Consider a set of T frequency-domain ordinates sampled, over the interval $[-\pi, \pi)$, from a boxcar function, centred on $\omega_0 = 0$. If the cut-off points lie between $\pm\omega_d = \pm 2\pi d/T$ and the succeeding Fourier frequencies, then the

sampled ordinates will be

$$\lambda_j = \begin{cases} 1, & \text{if } j \in \{-d, \ldots, d), \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

The corresponding filter coefficients will be given by

$$\beta^\circ(k) = \begin{cases} \dfrac{2d+1}{T}, & \text{if } k = 0, \\ \dfrac{\sin([d+1/2]\omega_1 k)}{T\sin(\omega_1 k/2)}, & \text{for } k = 1, \ldots, [T/2], \end{cases} \tag{31}$$

where $\omega_1 = 2\pi/T$. This function is just an instance of the Dirichlet kernel—see Pollock (1999), for example. Figure 1 depicts the frequency response for this filter at the Fourier frequencies, where $\lambda_j = 0, 1$. It also depicts the continuous frequency response that would be the consequence of applying an ordinary filter with these coefficients to a doubly-infinite data sequence.

It is notable that a casual inspection of the continuous response would lead one to infer the danger of substantial leakage, whereby elements that lie within the stop band are allowed to pass into the filtered sequence. In fact, with regard to the finite sample, there is no leakage.

If the cut-off points fall on the Fourier frequencies $\pm\omega_d$, and if $\lambda_d = \lambda_{-d} = 1/2$, then the filter coefficients will be

$$\beta^\circ(k) = \begin{cases} \dfrac{2d}{T}, & \text{if } k = 0, \\ \dfrac{\cos(\omega_1 k/2)\sin(d\omega_1 k)}{T\sin(\omega_1 k/2)}, & \text{for } k = 1, \ldots, [T/2]. \end{cases} \tag{32}$$

4. POLYNOMIALS AND THE MATRIX DIFFERENCE OPERATOR

The remaining sections of this paper are devoted to methods of filtering non-stationary sequences generated by linear stochastic processes with unit roots in an autoregressive operator. To support the analysis, the present section deals with the algebra of the difference operator and of the summation operator, which is its inverse. When it is used to cumulate data, the summation operator automatically generates polynomial functions of the discrete-time index. This feature is also analysed.

The matrix that takes the p-th difference of a vector of order T is

$$\nabla_T^p = (I - L_T)^p. \tag{33}$$

We may partition this matrix so that $\nabla_T^p = [Q_*, Q]'$, where Q_*' has p rows. If y is a vector of T elements, then

$$\nabla_T^p y = \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} y = \begin{bmatrix} g_* \\ g \end{bmatrix}; \tag{34}$$

and g_* is liable to be discarded, whereas g will be regarded as the vector of the p-th differences of the data.

The inverse matrix is partitioned conformably to give $\nabla_T^{-p} = [S_*, S]$. It follows that

$$[\, S_* \quad S\,] \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} = S_*Q_*' + SQ' = I_T, \tag{35}$$

and that

$$\begin{bmatrix} Q_*' \\ Q' \end{bmatrix} [\, S_* \quad S\,] = \begin{bmatrix} Q_*'S_* & Q_*'S \\ Q'S_* & Q'S \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_{T-p} \end{bmatrix}. \tag{36}$$

If g_* is available, then y can be recovered from g via

$$y = S_*g_* + Sg. \tag{37}$$

The lower-triangular Toeplitz matrix $\nabla_T^{-p} = [S_*, S]$ is completely characterised by its leading column. The elements of that column are the ordinates of a polynomial of degree $p-1$, of which the argument is the row index $t = 0, 1, \ldots, T-1$. Moreover, the leading p columns of the matrix ∇_T^{-p}, which constitute the submatrix S_*, provide a basis for all polynomials of degree $p-1$ that are defined on the integer points $t = 0, 1, \ldots, T-1$.

It follows that $S_*g_* = S_*Q_*'y$ contains the ordinates of a polynomial of degree $p-1$, which is interpolated through the first p elements of y, indexed by $t = 0, 1, \ldots, p-1$, and which is extrapolated over the remaining integers $t = p, p+1, \ldots, T-1$.

A polynomial that is designed to fit the data should take account of all of the observations in y. Imagine, therefore, that $y = \phi + \eta$, where ϕ contains the ordinates of a polynomial of degree $p-1$ and η is a disturbance term with $E(\eta) = 0$ and $D(\eta) = \Omega_\eta$. Then, in forming an estimate $f = S_*r_*$ of ϕ, we should minimise the sum of squares $\eta'\Omega_\eta^{-1}\eta$. Since the polynomial is fully determined by the elements of a starting-value vector r_*, this is a matter of minimising

$$(y-f)'\Omega_\eta^{-1}(y-f) = (y - S_*r_*)'\Omega_\eta^{-1}(y - S_*r_*) \tag{38}$$

with respect to r_*. The resulting values are

$$r_* = (S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1}y \quad \text{and} \quad f = S_*(S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1}y. \tag{39}$$

An alternative representation of the estimated polynomial is available, which avoids the inversion of Ω_η. This is provided by the identity

$$\begin{aligned} P_* &= S_*(S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1} \\ &= I - \Omega_\eta Q(Q'\Omega_\eta Q)^{-1}Q' = I - P_Q, \end{aligned} \tag{40}$$

which gives two representations of the projection matrix P_*. The equality follows from the fact that, if $\mathrm{Rank}[R, S_*] = T$ and if $S_*'\Omega_\eta^{-1}R = 0$, then

$$S_*(S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1} = I - R(R'\Omega_\eta^{-1}R)^{-1}R'\Omega_\eta^{-1}. \tag{41}$$

Setting $R = \Omega_\eta Q$ gives the result.

Grenander and Rosenblatt (1957) have shown that, when the disturbances are generated by a stationary stochastic process, the generalised least-squares (GLS) estimator of a polynomial regression that incorporates the matrix Ω_η is asymptotically equivalent to the ordinary least-squares (OLS) estimator that has the identity matrix I_T in place of Ω_η. The result has also been demonstrated by Anderson (1971). It means that one can use the OLS estimator without significant detriment to the quality of the estimates.

The result is proved by showing that the metric of the GLS regression is asymptotically equivalent to the Euclidean metric of OLS regression. It is required that, in the limit, the columns of $\Omega_\eta^{-1}S_*$ and S_* should span the same space. It follows that the projection operator P_* of (40) is asymptotically equivalent to $S_*(S_*'S_*)^{-1}S_*' = I - Q(Q'Q)^{-1}Q'$.

The residuals of an OLS polynomial regression of degree p, which are given by $y - f = Q(Q'Q)^{-1}Q'y$, contain same the information as the vector $g = Q'y$ of the p-th differences of the data. The difference operator has the effect of nullifying the element of zero frequency and of attenuating radically the adjacent low-frequency elements. Therefore, the low-frequency spectral structures of the data are not perceptible in the periodogram of the differenced sequence.

On the other hand, the periodogram of a trended sequence is liable to be dominated by its low-frequency components, which will mask the other spectral structures. However, the periodogram of the polynomial regression residuals can be relied upon to reveal the spectral structures at all frequencies. Moreover, by varying the degree p of the polynomial, one is able to alter the relative emphasis that is given to high-frequency and low-frequency structures.

5. WIENER–KOLMOGOROV FILTERING OF NONSTATIONARY DATA

The treatment of trended data must accommodate stochastic processes with drift. Therefore, it will be assumed that, within $y = \xi + \eta$, the trend component $\xi = \phi + \zeta$ is the sum of a vector ϕ, containing ordinates sampled from a polynomial in t of degree p at most, and a vector ζ from a stochastic process with p unit roots that is driven by a zero-mean process.

If Q' is the p-th difference operator, then $Q'\phi = \mu\iota$, with $\iota = [1, 1, \ldots, 1]'$, will contain a constant sequence of values, which will be zeros if the degree of the stochastic drift is less than p, which is the degree of differencing. Also, $Q'\zeta$ will be a vector sampled from a mean-zero stationary process. Therefore, $\delta = Q'\xi$ is from a stationary process with a constant mean. Thus, there is

$$\begin{aligned} Q'y &= Q'\xi + Q'\eta \\ &= \delta + \kappa = g, \end{aligned} \tag{42}$$

where

$$\begin{aligned} &E(\delta) = \mu\iota, \qquad D(\delta) = \Omega_\delta, \\ &E(\kappa) = 0, \qquad D(\kappa) = \Omega_\kappa = Q'\Omega_\eta Q, \\ &\text{and} \quad C(\delta, \kappa) = 0. \end{aligned} \tag{43}$$

Let the estimates of ξ, η, $\delta = Q'\xi$ and $\kappa = Q'\eta$ be denoted by x, h, d and k respectively. Then, with $E(g) = E(\delta) = \mu\iota$, there is

$$\begin{aligned} E(\delta|g) &= E(\delta) + \Omega_\delta(\Omega_\delta + \Omega_\kappa)^{-1}\{g - E(g)\} \\ &= \mu\iota + \Omega_\delta(\Omega_\delta + Q'\Omega_\eta Q)^{-1}\{g - \mu\iota\} = d, \end{aligned} \tag{44}$$

$$\begin{aligned} E(\kappa|g) &= E(\kappa) + \Omega_\kappa(\Omega_\delta + \Omega_\kappa)^{-1}\{g - E(g)\} \\ &= Q'\Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}\{g - \mu\iota\} = k; \end{aligned} \tag{45}$$

and these vectors obey an adding-up condition:

$$Q'y = d + k = g. \tag{46}$$

In (44), the lowpass filter matrix $Z_\delta = \Omega_\delta(\Omega_\delta + Q'\Omega_\eta Q)^{-1}$ will virtually conserve the vector $\mu\iota$, which is an element of zero frequency. In (45), the complementary highpass filter matrix $Z_\kappa = Q'\Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}$ will virtually nullify the vector. Its failure to do so completely is attributable the fact that the filter matrix is of full rank. As the matrix converges on its asymptotic form, the nullification will become complete. It follows that, even when the degree of the stochastic drift is p, one can set

$$d = \Omega_\delta(\Omega_\delta + \Omega_\kappa)^{-1}g = \Omega_\delta(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y, \tag{47}$$

$$k = \Omega_\kappa(\Omega_\delta + \Omega_\kappa)^{-1}g = Q'\Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{48}$$

The estimates of ξ and η may be obtained by integrating, or re-inflating, the components of the differenced data to give

$$x = S_*d_* + Sd \qquad \text{and} \qquad h = S_*k_* + Sk. \tag{49}$$

For this, the starting values d_* and h_* are required. The initial conditions in d_* should be chosen so as to ensure that the estimated trend is aligned as closely as possible with the data. The criterion is

$$\text{Minimise} \quad (y - S_*d_* - Sd)'\Omega_\eta^{-1}(y - S_*d_* - Sd) \quad \text{with respect to} \quad d_*. \tag{50}$$

The solution for the starting values is

$$d_* = (S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1}(y - Sd). \tag{51}$$

The starting values of k_* are obtained by minimising the (generalised) sum of squares of the fluctuations:

$$\text{Minimise} \quad (S_*k_* + Sk)'\Omega_\eta^{-1}(S_*k_* + Sk) \quad \text{with respect to} \quad k_*. \tag{52}$$

The solution is

$$k_* = -(S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1}Sk. \tag{53}$$

It is straightforward to show that, when (42) holds,

$$y = x + h, \tag{54}$$

which is to say that the addition of the estimates produces the original data series. As in (40), let $P_* = S_*\Omega_\eta^{-1}(S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'$. Then, the two components can be written as

$$x = Sd + S_*d_* = Sd + P_*(y - Sd), \tag{55}$$

$$h = Sk + S_*k_* = (I - P_*)Sk. \tag{56}$$

Adding these, using $d + k = Q'y$ from (46), gives

$$\begin{aligned} P_*y + (I - P_*)SQ'y &= P_*y + (I - P_*)(SQ' + S_*Q_*')y \\ &= P_*y + (I - P_*)y = y. \end{aligned} \tag{57}$$

Here, the first equality follows from the fact that $(I-P_*)S_* = 0$, and the second follows from the identity $SQ' + S_*Q_*' = I$, which is from (35). In view of (49), it can be seen that the condition $x + h = y$ signifies that the criteria of (50) and (52) are equivalent.

The starting values k_* and d_* can be eliminated from the expressions for x and h, which provide the estimates of the components. Substituting the expression for P_Q from (40) into $h = (I - P_*)Sk = P_QSk$ together with the expression for k from (48) and using the identity $Q'S = I_T$ gives

$$h = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{58}$$

A similar reduction can be pursued in respect of the equation $x = (I-P_*)Sd+P_*y = P_QSd + (I - P_Q)y$. However, it follows immediately from (54) that

$$\begin{aligned} x &= y - h \\ &= y - \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \end{aligned} \tag{59}$$

Observe that the filter matrix $Z_\eta = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}$ of (58), which delivers $h = Z_\eta g$, differs from the matrix $Z_\kappa = Q'Z_\eta$ of (48), which delivers $k = Z_\kappa g$, only in respect of the matrix difference operator Q'. The effect of omitting the operator is to remove the need for re-inflating the filtered components and thus to remove the need for the starting values.

Equation (59) can also be derived via a straightforward generalisation of the chi-square criterion of (10). If we regard the elements of δ_* as fixed values, then the dispersion matrix of $\xi = S_*\delta_* + S\delta$ is the singular matrix $D(\xi) = \Omega_\xi = S\Omega_\delta S'$. On setting $\eta = y - \xi$ in (10) and replacing the inverse of Ω_ξ^{-1} by the generalised inverse $\Omega_\xi^+ = Q\Omega_\delta^{-1}Q'$, we get the function

$$S(\xi) = (y - \xi)'\Omega_\eta^{-1}(y - \xi) + \xi'Q\Omega_\delta^{-1}Q'\xi, \tag{60}$$

of which the minimising value is

$$x = (Q\Omega_\delta^{-1}Q' + \Omega_\eta^{-1})^{-1}\Omega_\eta^{-1}y. \tag{61}$$

The matrix inversion lemma gives

$$(Q\Omega_\delta^{-1}Q' + \Omega_\eta^{-1})^{-1} = \Omega_\eta - \Omega_\eta Q(Q'\Omega_\eta Q + \Omega_\delta)^{-1}Q'\Omega_\eta; \tag{62}$$

and putting this into (61) gives the expression under (59). The matrix of (62) also constitutes the error dispersion matrix $D(\eta|y) = D(\xi|y)$ which, in view of their adding-up property (54), is common to the estimates of the two components.

It is straightforward to compute h when Ω_η and Ω_δ have narrow-band moving-average forms. Consider $h = \Omega_\eta Q b$, where $b = (Q'\Omega_\eta Q + \Omega_\delta)^{-1} g$. First b may be calculated by solving the equation $(Q'\Omega_\eta Q + \Omega_\delta)b = g$. This involves the Cholesky decomposition of $Q'\Omega_\eta Q + \Omega_\delta = GG'$, where G is a lower-triangular matrix with a number of nonzero diagonal bands equal to the recursive order of the filter. The system $GG'b = Gp = g$ is solved for p. Then, $G'b = p$ is solved for b. These are the recursive operations. Given b, then h can be obtained via direct multiplications.

6. APPLYING THE FOURIER METHOD TO NON-STATIONARY SEQUENCES

The method of processing nonstationary sequences that has been expounded in the preceding section depends upon the finite-sample versions of the difference operator and the summation operator. These are derived by replacing z in the functions $\nabla(z) = 1 - z$ and $\Sigma(z) = (1 - z)^{-1} = \{1 + z + z^2 + \cdots\}$ by the finite-sample lag operator L_T, which gives rise to a nonsingular matrix $\nabla(L_T) = [Q_*, Q]'$ and its inverse $\nabla^{-1}(L_T) = \Sigma(L_T) = [S_*, S]$.

The identity of (40), which is due to the interaction of the matrices S_* and Q', has enabled us to re-inflate the components of the differenced sequence $Q'y = g = d + k$ without explicitly representing the initial conditions or start-values d_* and k_* that are to be found in equations (54) and (55).

The operators that are derived by replacing of z in $\nabla(z)$ and $\Sigma(z)$ by the circulant lag operator K_T are problematic. On the one hand, putting K_T in place of z within $\nabla(z)$ results in a non-invertible matrix of rank $T - 1$. On the other hand, the replacement of z within the series expansion of $\Sigma(z)$ results in a non-convergent series that corresponds to a matrix in which all of the elements are uniformly of infinite value.

In applying the Fourier method of signal extraction to non-stationary sequences, one recourse is to use $\nabla(L_T)$ and $\Sigma(L_T)$ for the purposes of reducing the data to stationarity and for re-inflating them. The components δ and κ of the differenced data may estimated via the equations

$$d = \bar{U}\Lambda_\delta(\Lambda_\delta + \Lambda_\kappa)^{+}Ug = P_\delta g, \tag{63}$$

$$k = \bar{U}\Lambda_\kappa(\Lambda_\delta + \Lambda_\kappa)^{+}Ug = P_\kappa g. \tag{64}$$

For a vector $\mu\iota$ of repeated elements, there will be $P_\delta \mu\iota = \mu\iota$ and $P_\kappa \mu\iota = 0$.

Whereas the estimates of δ, κ may be extracted from $g = Q'y$ by the Fourier methods, the corresponding estimates x, h of ξ, η will be found from equations (55) and (56), which originate in the time-domain approach and which require explicit initial conditions.

It may also be appropriate, in this context, to replace the criteria of (50) and (52), which generate the values of d_* and k_*, by simplified criteria wherein

Ω_η is replaced by the identity matrix I_T. Then,

$$d_* = (S_*'S_*)^{-1}S_*'(y - Sd) \quad \text{and} \quad k_* = -(S_*'S_*)^{-1}S_*'Sk. \tag{65}$$

The available formulae for the summation of sequences provide convenient expressions for the values of the elements of $S_*'S_*$. (See, for example, Banerjee *et al.* 1993, p. 20.)

An alternative recourse, which is available in the case of a highpass or bandpass filter that nullifies the low-frequency components of the data, entails removing the implicit differencing operator from the filter. (In an appendix of their paper, Baxter and King (1999) demonstrate the presence, within a symmetric bandpass filter, of two unit roots, i.e. of a twofold differencing operator.)

Consider a filter defined in respect of a doubly-infinite sequence, and let $\phi(z)$ be the transfer function of the filter, i.e. the z-transform the filter coefficients. Imagine that $\phi(z)$ contains the factor $(1 - z)^p$, and let $\psi(z) = (1-z)^{-p}\phi(z)$. Then, $\psi(z)$ defines a filter of which the finite-sample version can be realised by the replacement of z by K_T.

Since $K_T = \bar{U}DU$, the filter matrix can be factorised as $\psi(K_T) = \Psi = \bar{U}\psi(K_T)U$. On defining $J_\psi = \psi(K_T)$, which is a diagonal weighting matrix, the estimate of the signal component is given by the equation

$$x = \bar{U}J_\psi U g. \tag{66}$$

Whichever procedure is adopted, the Fourier method is applied to data that have been reduced to stationarity by differencing. Given the circularity of the method, if follows that, as a filter approaches one end of the sample it will take an increasing proportion of its data from the other end. To avoid some of the effects of this, one may extrapolate the data at both ends of sample. The extrapolation can be done either before or after the data has been differenced. After the filtering, the extra-sample elements may be discarded.

If the trended data has been extrapolated, then the starting values d_* and k_*, which are required by the first procedure, may be obtained by minimising the quadratic functions (50) and (52) defined over the enlarged data set. Otherwise, if the differenced data have been extrapolated, then the extra-sample points of the filtered sequence are discarded prior to its re-inflation, and the quadratic functions that are to be minimised in pursuit of the starting values are defined over T points corresponding to the original sample.

Extrapolating the differenced data has the advantage of facilitating the process of tapering whereby the ends of the sample are reduced to a level at which they can meet when the data are wrapped around a circle. The purpose of this is to avoid any radical disjunctions that might otherwise occur at the meeting point and which might have ill effects throughout the filtered sequences.

We should also note that, in the case of a finite bandpass filter, a third possibility arises. This is to apply the filter directly to the original, undifferenced, data. Such a procedure has been followed both by Baxter and King (1999) and by Christiano and Fitzgerald (2003). The latter have found an ingenious way of extrapolating the data, so as to enable the filter to generate values at the ends of the sample.

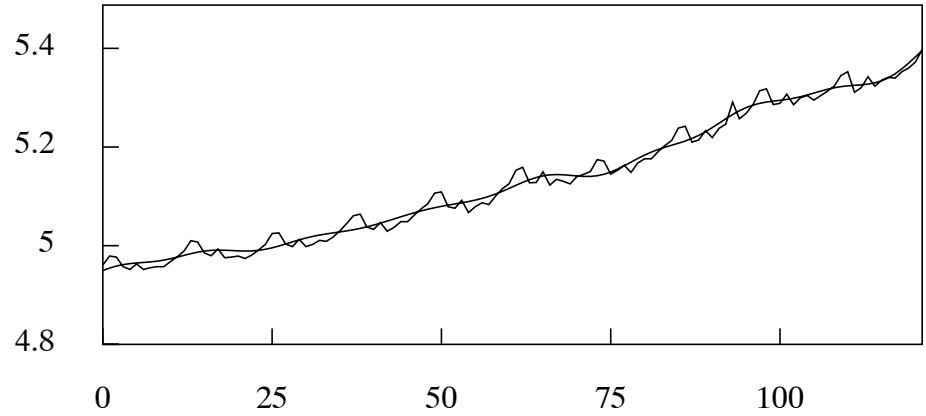

Figure 2. The plot of the logarithms of 132 monthly observations on the U.S. money supply, beginning in January 1960. A trend, estimated by the Fourier method, has been interpolated through the data.

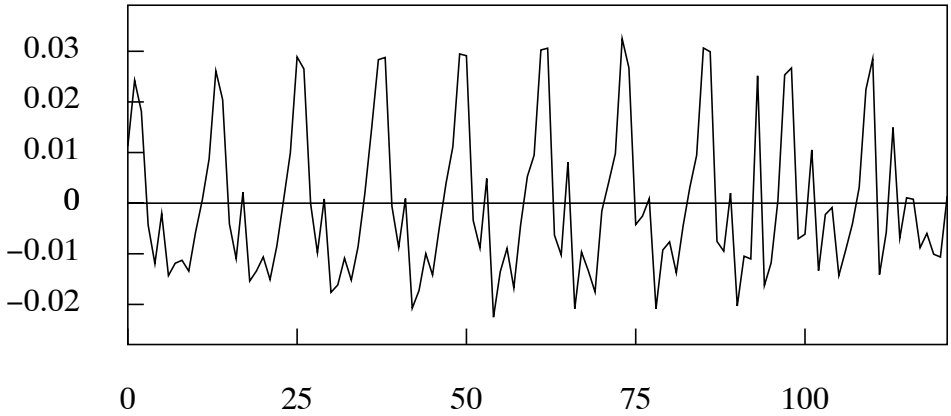

Figure 3. The deviations of the logarithmic money-supply data from the interpolated trend.

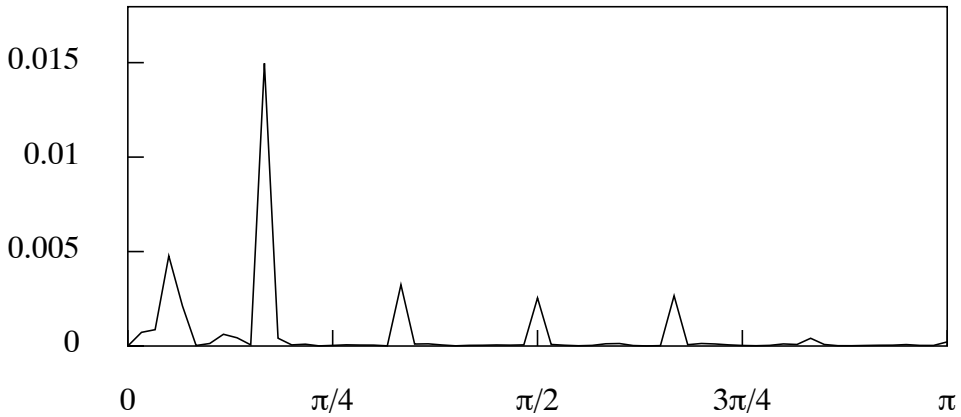

Figure 4. The periodogram of the residuals from the quadratic detrending of the logarithmic money-supply data.

The Fourier method can be used to advantage whenever the components of the data are to be found in disjoint frequency intervals. If the components are separated by wide spectral dead spaces, then it should be possible extract them from the data with filters that have gradual transitions that are confined to the dead spaces. In that case, Wiener–Kolmogorov time-domain filters of low orders can be used. However, when the components are adjacent or are separated by narrow dead spaces, it is difficult to devise a time-domain filter with a transition that is sharp enough and that does not suffer from problems of instability. Then, a Fourier method should be used instead.

Example. Figure 2 show the logarithms of a monthly sequence of 132 observations of the U.S. money supply, through which a trend has been interpolated, and Figure 3 shows the sequence of the deviations from the tend. The data are from Bolch and Huang (1974).

The trend has been estimated with reference to Figure 4, which shows the periodogram of the residuals from a quadratic detrending of the data. There is a tall spike in the periodogram at $\omega_{11} = \pi/6$, which represents the fundamental frequency of an annual seasonal fluctuation. To the left of this spike, there is a spectral mass, which belongs the trend component. Its presence is an indication of the inadequacy of the quadratic detrending. To remove this spectral mass and to estimate the trend more effectively, a Fourier method has been deployed.

An inspection of the numerical values of the ordinates of the periodogram has indicated that, within the accuracy of the calculations, the element at ω_{10}, which is adjacent to the seasonal frequency, is zero-valued. This defines the point of transition of the lowpass filter that has been used to isolate the trend component of the data. A filter with a less abrupt transition would allow the powerful seasonal component to leak into the trend, and vice versa. Experiments, using Wiener–Kolmogorov time-domain filters, have shown the detriment of this.

The trend in Figure 2 has been created by re-inflating a sequence that has been synthesised from the relevant ordinates selected from the Fourier transform of the twice differenced data. The latter have a positive mean value, which corresponds to a quadratic drift in the original data. The differenced data have not been extrapolated, since there appears to be no detriment in the end-effects. Moreover, the regularity of the seasonal fluctuations, as revealed by the sequence of the deviations from the trend in Figure 3, is a testimony to the success in estimating the trend.

REFERENCES

Anderson, T.W., (1971), *The Statistical Analysis of Time Series.* John Wiley and Sons, New York.

Banerjee, A., J. Dolado, J.W. Galbraith and D.F. Hendry, (1993), *Co-integration, Error-Correction and the Econometric Analysis of Non-Stationary Data.* Oxford University Press, Oxford.

Baxter, M., and R.G. King, (1999), Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *Review of Economics and Statistics* 81, 575–593.

Bolch, B.W., and C.J. Huang, (1974), *Multivariate Statistical Methods for Business and Economics.* Prentice-Hall, Englewood Cliffs, N.J.

Christiano, L.J., and T.J. Fitzgerald, (2003). The Band-Pass Filter. *International Economic Review* 44, 435–465

Grenander, U., and M. Rosenblatt, (1957), *Statistical Analysis of Stationary Time Series.* John Wiley and Sons, New York.

Gray, R.M., (2002), *Toeplitz and Circulant Matrices: A Review.* Information Systems Laboratory, Department of Electrical Engineering, Stanford University, California, `(http://ee.stanford.edu/gray/~toeplitz.pdf).`

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

Pollock, D.S.G., (2000), Trend Estimation and De-trending via Rational Square Wave Filters. *Journal of Econometrics* 99, 317–334.

Pollock, D.S.G., (2002), Circulant Matrices and Time-Series Analysis. *The International Journal of Mathematical Education in Science and Technology* 33, 213–230.

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