METHODOLOGY FOR TREND ESTIMATION

by D.S.G. Pollock

Queen Mary and Westfield College

University of London

This paper describes a methodology for trend estimation which relies upon the finite-sample implementation of the classical Wiener–Kolmogorov theory of signal extraction in which provisions are made for dealing with a nonstationary signal component. It is argued that de-trending filters should be selected primarily on the basis of their frequency-response characteristics.

1. Introduction

The problem of trend estimation in econometrics has had a long history, and the techniques which can be deployed have been evolving gradually over many years. The forces of evolution have been twofold. On one hand is the gradual improvement in statistical and computational techniques which has been accompanied by improvements in the processing power of computers and in the accessibility of software programs. On the other hand are the methodological developments within the discipline of econometrics.

The econometric approach to trend estimation is based upon the notion that a time series is composed of several components of independent origin which are combined by addition or by multiplication. Usually, a multiplicative combination can be reduced to an additive one by the simple expedient of taking logarithms.

The components of the time series can be regarded as Fourier combinations of trigonometrical functions—i.e. of sines and cosines—whose frequencies fall within specified ranges. Over the range of the frequencies which pertain to a particular component, one can define a spectral density function which represents the squared amplitudes of the constituent trigonometrical functions.

If the frequency ranges of the components are completely segregated, then it is possible, in principle, to achieve a definitive separation of the time series into its independent components. If the frequency ranges of the components overlap, then it is possible to achieve a tentative separation in which trigonometrical functions of the same frequencies are present in two or more components of the time series. In each estimated component, the functions acquire the amplitudes which are indicated by the appropriate spectral density function. Since the estimates of such overlapping components originate from the same empirical data sequence, they are bound to be correlated with each other—which contradicts the theoretical assumptions regarding the components.

In recent years, the mainstay of econometric trend estimation has been the Wiener–Kolmogorov theory of signal extraction (See [10] and [17]). The theory

1

shows how to construct linear filters which preserve components of certain frequencies and which attenuate or nullify components at other frequencies. The effects of such filters are commonly described in terms of metaphors which borrow concepts from the physics of sound and light. Filtering a data series is akin to filtering light through a coloured lens.

The Wiener–Kolmogorov theory was developed for purposes which were quite different from those envisaged in econometric trend estimation. It was applied originally to stationary signals on which the observations are so abundant that they can be treated as if they constitute series which stretch indefinitely in both directions from any chosen point in time.

Econometric series are, usually, of a strictly limited duration, and, often, they manifest strong trends. It has been recognised, for some time, that trended or nonstationary sequences present no essential difficulties to the theory of linear filtering. In particular, if the observable time series is the sum of a nonstationary signal component and a stationary noise component, then the signal and the noise can be separated with no more difficulty than in the case of a stationary signal. See, for example, Pierce [13] and Bell [1].

The problems which are due to the limited durations of economic series are more difficult to handle. One approach has been to extend the data set in both directions using forecast and backcast values. By providing a set of plausible pre-sample values, one can stabilise the filter in the run-up to the data series so that its output of processed values is not seriously affected by the problem of the initial conditions. By providing post-sample values which represent a plausible future, one can facilitate the reverse-time filtering which is associated with phase-neutral methods which employ rational infinite-impulse-response filters—which are feedback filters in other words. An example of this kind of bidirectional filtering involving feedback was provided by Burman [3].

In this paper, we shall present a solution to the start-up problem which does not require any extra-sample values. The solution appears to be a definitive one. Other, similar, solutions which have been proposed recently have made use of sophisticated variants of the Kalman filter and of the associated smoothing algorithms. Thus, for example, Koopman, Harvey et al. [11] have used the diffuse Kalman filter algorithm of De Jong [5], whilst Gómez and Maravall [6] have proposed another adaptation of the Kalman filter. These algorithms are complicated, and it may be fair to suggest that they are fully understood only by a small group of analysts.

The difficulties are due to the general and all-encompassing nature of the algorithms. The simpler approach which we offer here deals in terms of the specific features of the problem at hand. Our approach can be depicted as a special case of Kalman filtering. It can also be assimilated to another branch of mathematics which deals with problems of smoothing and graduation which has found extensive application in industrial design via the Reinsch smoothing

2

spline [15].

In this paper, we shall examine three distinct approaches to the estimation of econometric trends. The first approach rests upon the model-based method of seasonal adjustment which has been advocated by Hillmer and Tiao [9]. This entails the so-called canonical decomposition of a seasonal ARIMA model. The second approach rests upon the so-called structural time-series model which has been proposed by Harvey and Todd [8] and which has been expounded at length by Harvey in a book [7]. The third approach is based upon a model which suppresses the seasonal component which is present in the two previous approaches.

We shall also draw attention to certain problems with arise from the manner in which a seasonal time series is usually modelled by placing complex roots of unit modulus in the autoregressive component of an ARMA model.

2. Filtering Nonstationary Sequences

The observable sequence, which is a function mapping from the set of integers $\mathcal{I} = \{t = 0, \pm 1, \pm 2, \ldots\}$ onto the real line, may be represented by

$$y(t) = \xi(t) + \eta(t). \tag{1}$$

This consists of two components which are assumed to be mutually uncorrelated. The first of these, which is the trend component, is modelled by an autoregressive integrated moving-average (ARIMA) process which can be written as

$$\xi(t) = \frac{\mu(L)}{\nabla^d(L)\alpha(L)} \nu(t). \tag{2}$$

Here $\alpha(L)$ and $\mu(L)$ are, respectively, an autoregressive and a moving-average operator with roots which lie outside the unit circle, whilst $\nabla^d(L) = (1 - L)^d$ is the dth power of the difference operator. The sequence $\nu(t)$ is generated by a white-noise process with $V\{\nu(t)\} = \sigma_\nu^2$.

The second component of the observable sequence is the residue which is represented by a seasonal autoregressive moving-average (SARMA) model in the form of

$$\eta(t) = \frac{\theta(L)}{S(L)\phi(L)} \varepsilon(t). \tag{3}$$

Here $\phi(L)$ and $\theta(L)$ are, respectively, an autoregressive and a moving-average operator with roots which lie outside the unit circle, whilst $S(L) = 1 + L + \cdots + L^{s-1}$ stands for a polynomial in the lag operator which has complex roots of unit modulus whose arguments correspond to a seasonal frequency and to

3

its harmonics. The sequence $\varepsilon(t)$ is generated by a white-noise process with $V\{\varepsilon(t)\} = \sigma_\varepsilon^2$.

The residual component $\eta(t)$ lumps together all the elements which are not comprised by the trend. In a structural model, $\eta(t)$ is presented as a sum of statistically independent components which may include a cyclical component, a seasonal component and an irregular white-noise component. If each of the components is represented by an ARMA process—which may have autoregressive roots of unit modulus—then the sum must also be an ARMA process taking the form of (3).

Consider the operator

$$(4) \qquad \delta(L) = \nabla^d(L)S(L)\alpha(L)\phi(L),$$

which is the product of the denominators of (2) and (3). This is a polynomial function of the lag operator of a degree which will be denoted by p. Multiplying $y(t)$ by $\delta(L)$ reduces it to a stationary process which is a sum of moving-average components. This can be written as

$$(5) \qquad q(t) = \zeta(t) + \kappa(t),$$

where

$$(6) \qquad q(t) = \delta(L)y(t) = \delta(L)\xi(t) + \delta(L)\eta(t),$$

$$(7) \qquad \zeta(t) = \delta(L)\xi(t) = \psi_T(L)\nu(t), \quad \text{and}$$

$$(8) \qquad \kappa(t) = \delta(L)\eta(t) = \psi_R(L)\varepsilon(t).$$

Here we have defined

$$(9) \qquad \psi_T(L) = S(L)\phi(L)\mu(L) \qquad \text{and} \qquad \psi_R(L) = \nabla^d(L)\alpha(L)\theta(L).$$

On substituting (7) and (8) into (5) we obtain

$$(10) \qquad q(t) = \psi_T(L)\nu(t) + \psi_R(L)\varepsilon(t),$$

which is an expression which we shall have occasion to use later.

Our objective is to obtain a decomposition

$$(11) \qquad y(t) = x(t) + h(t),$$

in which

$$(12) \qquad x(t) = E\{\xi(t)|y(t)\} \qquad \text{and} \qquad h(t) = E\{\eta(t)|y(t)\}$$

are estimates of the trend component and of the residue, respectively.

At first sight, this purpose seems to be hindered by the fact that, in general, the sequences in (11) are nonstationary; for the Wiener–Kolmogorov theory of signal extraction relates to stationary processes. However, it is straightforward to show that the filter which serves to estimate the components $\zeta(t)$ and $\kappa(t)$ from the stationary sequence $q(t) = \delta(L)y(t)$ can be applied equally to the nonstationary sequence $y(t)$ in pursuance of the estimates of the corresponding components $\xi(t)$ and $\eta(t)$.

Our immediate object, however, is to devise a linear filter $\beta_T(L)$ which can be applied to the stationary sequence $q(t)$ to obtain an estimate of $\zeta(t)$ in the form of

$$(13) \qquad z(t) = E\big\{\zeta(t)|q(t)\big\} = \beta_T(L)q(t).$$

The complementary filter $\beta_R(L) = 1 - \beta_T(L)$ can be applied likewise to $q(t)$ to obtain an estimate of $\kappa(t)$ in the form of

$$(14) \qquad k(t) = E\big\{\kappa(t)|q(t)\big\} = \beta_R(L)q(t).$$

Thus the aim is to decompose $q(t) = \delta(L)y(t)$ as

$$(15) \qquad q(t) = z(t) + k(t).$$

From such estimates, we can recover the estimates $x(t) = \delta^{-1}(L)z(t)$ and $h(t) = \delta^{-1}(L)k(t)$ of the nonstationary signal or trend sequence $\xi(t)$ and of the residual sequence $\eta(t)$.

3. Minimum-Mean-Square-Error Filters

The coefficients of the optimal linear signal-extraction filter $\beta_T(L) = \sum_j \beta_j L^j$ are estimated by invoking the minimum-mean-square-error criterion. The errors in question are the elements of the sequence $e(t) = \zeta(t) - z(t)$, where $z(t)$ is given by equation (13). The principle of orthogonality, by which the criterion is fulfilled, indicates that the errors must be uncorrelated with the elements in the information set $\mathcal{I}_t = \{q_{t-k}; k = 0, \pm 1, \pm 2, \ldots\}$. Thus

$$
(16) \qquad
\begin{aligned}
0 &= E\Big\{ q_{t-k}(\zeta_t - z_t) \Big\} \\
&= E(q_{t-k}\zeta_t) - \sum_j \beta_j E(q_{t-k}q_{t-j}) \\
&= \gamma_k^{q\zeta} - \sum_j \beta_j \gamma_{k-j}^{qq},
\end{aligned}
$$

for all k. The equation may be expressed, in terms of the z transform, as

$$(17) \qquad \gamma^{q\zeta}(z) = \gamma^{qq}(z)\beta_T(z),$$

where $\beta_T(z)$ stands for an indefinite two-sided Laurent series comprising both positive and negative powers of z.

Given the assumption that the elements of the noise sequence $\kappa(t)$ are independent of those of the signal $\zeta(t)$, it follows that

$$(18) \qquad \gamma^{qq}(z) = \gamma^{\zeta\zeta}(z) + \gamma^{\kappa\kappa}(z) \qquad \text{and} \qquad \gamma^{q\zeta}(z) = \gamma^{\zeta\zeta}(z),$$

where

$$(19) \qquad \gamma^{\zeta\zeta}(z) = \sigma_\nu^2 \psi_T(z)\psi_T(z^{-1}) \qquad \text{and} \qquad \gamma^{\kappa\kappa}(z) = \sigma_\varepsilon^2 \psi_R(z)\psi_R(z^{-1}).$$

It follows from (17) that

$$(20) \qquad \beta_T(z) = \frac{\gamma^{\zeta\zeta}(z)}{\gamma^{qq}(z)} = \frac{\psi_T(z)\psi_T(z^{-1})}{\psi(z)\psi(z^{-1})},$$

where

$$(21) \quad \psi(z)\psi(z^{-1}) = \psi_T(z)\psi_T(z^{-1}) + \lambda\psi_R(z)\psi_R(z^{-1}) \qquad \text{with} \qquad \lambda = \frac{\sigma_\varepsilon^2}{\sigma_\nu^2}.$$

Here we have a sum of two positive definite functions which is itself a postive definite function. Therefore the sum can be factorised as $\psi(z)\psi(z^{-1})$.

Since, by assumption, the polynomials $\psi_T(z)$ and $\psi_R(z)$ have no roots of unit modulus in common, the polynomial $\psi(z)\psi(z^{-1})$ will not have any unit roots, which must be the case if the filter is to be stable and capable of implementation. Its roots will come in reciprocal pairs; and, once these are available, they can be assigned unequivocally to the factors $\psi(z)$ and $\psi(z^{-1})$. Those roots which lie outside the unit circle belong to $\psi(z)$ and those which lie inside belong to $\psi(z^{-1})$.

By setting $z = e^{i\omega}$, one can derive the frequency-response function of the filter which is used in estimating the signal $\zeta(t)$. The effect of the filter is to multiply each of the frequency components of $q(t)$ by the fraction of its variance which is attributable to the signal $\zeta(t)$. The same principle applies to the estimation of the noise or residue component $\kappa(t)$.

The residue-estimation filter is just the complementary filter

$$(22) \qquad \beta_R(z) = 1 - \beta_T(z) = \lambda\frac{\psi_R(z)\psi_R(z^{-1})}{\psi(z)\psi(z^{-1})}.$$

It will assist the subsequent exposition if we present more explicit expressions for the two filters. Thus, on substituting the expression for $\psi_T(z)$ from (9) into (20) we get,

$$(23) \qquad \beta_T(L) = S(L^{-1}) \left\{ \frac{\phi(L^{-1})\mu(L^{-1})\mu(L)\phi(L)}{\psi(L^{-1})\psi(L)} \right\} S(L).$$

Likewise, on substituting the expression for $\psi_R(z)$ from (9) into (22) we get

$$(24) \qquad \beta_R(L) = \lambda \nabla^d(L^{-1}) \left\{ \frac{\alpha(L^{-1})\theta(L^{-1})\theta(L)\alpha(L)}{\psi(L^{-1})\psi(L)} \right\} \nabla^d(L).$$

In summarising the development so far, we find that the formulae for estimating the components of the stationary sequence $q(t)$ of (5) are

$$(25) \qquad \begin{aligned} z(t) &= \beta_T(L)q(t) = \beta_T(L)\delta(L)y(t) \qquad \text{and} \\ k(t) &= \beta_R(L)q(t) = \beta_R(L)\delta(L)y(t). \end{aligned}$$

The error sequence which is associated equally with the estimates of $\zeta(t)$ and $\kappa(t)$ is given by

$$(26) \qquad \begin{aligned} e(t) &= \zeta(t) - z(t) \\ &= \psi_T(L)\nu(t) - \beta_T(L)\big\{\psi_T(L)\nu(t) + \psi_R(L)\varepsilon(t)\big\} \\ &= \beta_R(L)\psi_T(L)\nu(t) - \beta_T(L)\psi_R(L)\varepsilon(t). \end{aligned}$$

Here, the second equality derives from the expression $\zeta(t) = \psi_T(L)\nu(t)$ from (7) and the expression $z(t) = \beta_T(L)q(t)$ from (13). Into the latter expression, we must substitute the expression for $q(t)$ from (10).

From the estimates $z(t)$ and $k(t)$, we can recover the estimates $x(t)$ and $h(t)$ of the components of $y(t) = \xi(t) + \eta(t)$. The estimates are

$$(27) \qquad \begin{aligned} x(t) &= \delta^{-1}(L)z(t) = \beta_T(L)y(t) \qquad \text{and} \\ h(t) &= \delta^{-1}(L)k(t) = \beta_R(L)y(t). \end{aligned}$$

Thus $x(t)$ and $h(t)$ may be obtained by applying the filters directly to $y(t)$ rather than to its transformed version $q(t)$. Alternatively, if the filters are applied to the stationary sequence $q(t)$, then the estimates can be recovered from $z(t)$ and $k(t)$ by use of the inverse operator $\delta^{-1}(L)$; and there is liable to be some computational advantages in taking this approach. The reason is that the elements of $q(t)$ are bound to have a smaller numerical range that those of $y(t)$.

The error sequence which is associated equally with the estimates of the trend $\xi(t)$ and of the residue $\eta(t)$ is given by

(28)
$$\begin{aligned}
\delta^{-1}(L)e(t) &= \delta^{-1}(L)\zeta(t) - \delta^{-1}(L)z(t) \\
&= \delta^{-1}(L)\beta_R(L)\psi_T(L)\nu(t) - \delta^{-1}(L)\beta_T(L)\psi_R(L)\varepsilon(t).
\end{aligned}$$

It will be found that the factors $\nabla^d(L)$ and $S(L)$, which contain roots of unit modulus, can be eliminated from $\delta^{-1}(L)\beta_R(L)\psi_T(L)$ and $\delta^{-1}(L)\beta_T(L)\psi_R(L)$ by cancellations between the numerators and denominators of these operators. Thus the sequence $\delta^{-1}(L)e(t)$ of the cumulated errors is seen to be the product of a stationary process. This is, of course, a crucial outcome; for, were it not the case, then it would be incorrect to describe $x(t)$ and $h(t)$ as the minimum-mean-square-error estimates; and the estimates would be useless.

Given the complementary nature of the estimates $x(t)$ and $h(t)$, only one of them needs to be computed. The second component can be obtained by subtracting the first component from $y(t)$. Let us devote our effort to finding $h(t)$, for the reason that it is more likely to be stationary that is $x(t)$.

Reference to (24) shows that the computation of $h(t) = \beta_R(L)y(t)$ can be broken down into four stages:

(29)
$$d(t) = \nabla^d(L)y(t),$$

(30)
$$f(t) = \frac{\alpha(L)\theta(L)}{\psi(L)}d(t),$$

(31)
$$g(t) = \frac{\alpha(L^{-1})\theta(L^{-1})}{\psi(L^{-1})}f(t),$$

(32)
$$h(t) = \lambda\nabla^d(L^{-1})g(t).$$

We should note that the seasonal operator S, which seems to be missing from these expressions, is, in fact, buried within ψ.

The first stage (29) involves the successive differencing of the sequence $y(t)$. If there are no seasonal roots and the seasonal operator is, in fact, the identity operator $S = I$, then the differencing will be sufficient to reduce the sequence to stationarity. The second stage (30) represents a process of feedback filtering which runs in the direction of time. The filter $\alpha(L)\theta(L)/\psi(L)$ is stable in consequence of the condition that all of the roots of $\psi(z)$ lie outside the unit circle. The third stage (31) is the reverse of the second stage and it represents a process of feedback filtering which runs in reversed time. The final stage (32) is the reverse of the first stage.

We may observe that the filter $\beta_R(z)$ is phase neutral in the sense that it induces neither lead nor lags in the processed series $h(t)$. The same is true

of the trend-estimation filter $\beta_T(z)$. These results are the consequence of the symmetry of the filters. In effect, the reverse-time processes of (31) and (32) serve to eliminate any lags which may be induced by the processes (29) and (30) by inducing equal and opposite reverse-time lags.

The equations (29)–(32) presuppose that the sequences which are to be filtered are defined for all positive and negative integers. In practice, the sequences are finite and, in econometric applications, they are liable to be of strictly limited duration. Therefore there can be a serious mismatch between the assumptions from which the filters are derived and the circumstances in which they are applied.

One way of coping with the limitations of the observations on the data sequence $y(t)$ is to supplement them by pre-sample and post-sample values obtained by the normal methods of forecasting. Then 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, or with a plausible future, if it is working in reversed time. For short series, the quality of the estimates of the trend and the residue are liable to be heavily dependent upon the quality of these extrapolations which need to be close to the true values.

If both the trend and the residue were truly generated by nonstationary stochastic processes, then the errors of the extrapolations would also be non-stationary. In that case, the replacement of the extra-sample values by their forecasts would not be a viable option. Notice, however, that such problems do not arise when the seasonal operator S is absent from the process, depicted in equation (3), which generates $\eta(t)$. In that case, the sequence $\nabla^d(L)y(t)$, which is found in equation (29), will be stationary; and its pre-sample values may be represented by its zero-valued unconditional expectation. Likewise, the post-sample values of $f(t)$, which are required in the run-up to the reverse-time filtering process, depicted by equations (31) and (32), could also be represented by zeros.

In fact, the seasonal fluctuations which are present in econometric data series bear only a limited resemblance to the sort of nonstationary stochastic process depicted by equation (3). In particular, whenever the operator S contains unit roots, the quasi-cyclical process $\eta(t)$ will be theoretically unbounded. Also, the phase of its cycles will vary in a haphazard manner. By contrast, the seasonal fluctuations in economic data are closely bounded and almost constant in amplitude, and their peaks and troughs are likely to occur perennially in specific months of the year.

There is, however, a close resemblance between seasonal fluctuations in econometric series and the trajectories of the forecast functions of seasonal ARIMA models which tend towards perfectly regular cycles which are linear combinations of trigonometrical functions. The ability of seasonal ARIMA

models accurately to forecast the seasonal cycles ensures the viability, in practice, of the Wiener–Kolmogorov filters which are based on the formulations of structural time-series models.

4. Extracting Signals from Finite Samples

In this section, we shall develop an approach to trend estimation which deals explicitly with the fact that a data series is of a finite duration. It will be left largely to the reader to trace the manifest connections between this approach and the approach, pursued in the previous section, which begins by assuming that the data series is of an infinite duration.

Let us imagine, therefore, that there are only T observations of the process $y(t)$ of equation (1) which run from $t = 0$ to $t = T - 1$. These are gathered in a vector

$$(33) \qquad\qquad y = \xi + \eta.$$

To find the finite-sample the counterpart of equation (5), we need to represent the operator $\delta(L)$ of (4) in the form of a matrix. The matrix, which is of order $(T - p) \times T$ and which contains the full set of coefficients of the polynomial $\delta(z)$ in each successive row, is in the form of $\Delta' = [\Delta'_1, \Delta'_2]$, where

$$(34) \quad \Delta'_1 = \begin{bmatrix} \delta_p & \dots & \delta_1 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \delta_p \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \\ 0 & \dots & 0 \end{bmatrix}, \quad \Delta'_2 = \begin{bmatrix} 1 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ \delta_{p-1} & \dots & 1 & 0 & \dots & 0 & 0 \\ \delta_p & \dots & \delta_1 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & & & \ddots & \vdots & \vdots \\ 0 & \dots & \delta_p & \delta_{p-1} & \dots & 1 & 0 \\ 0 & \dots & 0 & \delta_p & \dots & \delta_1 & 1 \end{bmatrix}.$$

Observe that, if y were a vector of T values of a polynomial of degree $d - 1$, taken at equally-spaced intervals, then we should have $\Delta'y = 0$. Here, d is the order of the difference operator $\nabla^d(L)$, which is a factor of $\delta(L)$.

Premultiplying equation (33) by Δ' gives

$$(35) \qquad \begin{aligned} q = \Delta'y &= \Delta'\xi + \Delta'\eta \\ &= \zeta + \kappa, \end{aligned}$$

where $\zeta = \Delta'\xi$ and $\kappa = \Delta'\eta$. The first and second moments of the vector ζ may be denoted by

$$(36) \qquad E(\zeta) = 0 \quad \text{and} \quad D(\zeta) = \sigma_\nu^2 \Omega_T,$$

and those of κ by

$$(37) \qquad\qquad E(\kappa) = 0 \qquad \text{and} \qquad D(\kappa) = \sigma_\varepsilon^2 \Omega_R,$$

where Ω_T and Ω_R are symmetric Toeplitz matrices with a limited number of nonzero diagonal bands. The generating functions for the coefficients of these dispersion matrices are, respectively, the functions $\gamma^{\zeta\zeta}(z)$ and $\gamma^{\kappa\kappa}(z)$ of (19).

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

$$(38) \qquad \begin{aligned} E(\zeta|q) &= E(\zeta) + C(\zeta,q)D^{-1}(q)\{q - E(q)\} \\ &= \Omega_T(\Omega_T + \lambda\Omega_R)^{-1}q = z, \end{aligned}$$

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

$$(39) \qquad \begin{aligned} E(\kappa|q) &= E(\kappa) + C(\kappa,d)D^{-1}(q)\{q - E(q)\} \\ &= \lambda\Omega_R(\Omega_T + \lambda\Omega_R)^{-1}q = k. \end{aligned}$$

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

The estimates are calculated, first, by solving the equation

$$(40) \qquad\qquad (\Omega_T + \lambda\Omega_R)g = q$$

for the value of g and, thereafter, by finding

$$(41) \qquad\qquad z = \Omega_T g \qquad \text{and} \qquad k = \lambda\Omega_R g.$$

The solution of equation (40) is found via a Cholesky factorisation which sets $\Omega_T + \lambda\Omega_R = GG'$, where G is a lower-triangular matrix. The system $GG'g = q$ may be cast in the form of $Gh = q$ and solved for h. Then $G'g = h$ can be solved for g.

Our object is to recover from z an estimate x of the trend vector ξ. This would be conceived, ordinarily, as a matter of pursuing a simple recursion which is based upon the equation

$$(42) \qquad\qquad x_t = z_t - \delta_1 x_{t-1} - \cdots - \delta_p x_{t-p},$$

with the index t running from $t = 0$ to $t = T - 1$. The difficulty is in discovering the appropriate initial conditions $x_{-1}, \ldots, x_{-p}$ with which to begin the recursion.

We can circumvent the problem of the initial conditions by seeking the solution to the following problem:

$$(43) \qquad \text{Minimise} \qquad (y - x)'\Sigma^{-1}(y - x) \qquad \text{Subject to} \qquad \Delta'x = z,$$

where Σ is a positive definite matrix which defines an appropriate metric. This entails the minimisation of a generalised sum of square of residuals which are the deviations of the trend vector x from the data vector y. The constraint is that the transformed value $z = \Delta' x$ of the trend vector must equal the conditional expectation $z = E(\zeta|q)$ specified in (38).

If the process which generates η is stationary, which is to say that there are no seasonal unit roots in the denominator of equation (3), then η has a well-defined dispersion matrix, and we should have $D(\eta) = \sigma_\varepsilon^2 \Sigma$. In that case, (43) becomes a conventional generalised least-squares criterion.

The problem of (43) is addressed by evaluating the Lagrangean function

$$(44) \qquad L(x, \mu) = (y - x)'\Sigma^{-1}(y - x) + 2\mu'(\Delta'x - z).$$

We may describe this as the restricted least-squares criterion function.

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

$$(45) \qquad \Sigma^{-1}(y - x) - \Delta\mu = 0.$$

Premultiplying by $\Delta'\Sigma$ gives

$$(46) \qquad \Delta'(y - x) = \Delta'\Sigma\Delta\mu.$$

But, from (40) and (41), it follows that

$$(47) \qquad \begin{aligned} \Delta'(y - x) &= q - z \\ &= \lambda\Omega_R g; \end{aligned}$$

whence we get

$$(48) \qquad \begin{aligned} \mu &= (\Delta'\Sigma\Delta)^{-1}\Delta'(y - x) \\ &= \lambda(\Delta'\Sigma\Delta)^{-1}\Omega_R g. \end{aligned}$$

Putting the final expression for μ into (45) gives

$$(49) \qquad \begin{aligned} x &= y - \Sigma\Delta\mu \\ &= y - \lambda\Sigma\Delta(\Delta'\Sigma\Delta)^{-1}\Omega_R g. \end{aligned}$$

This is our solution to the problem of estimating the trend vector ξ. In the case where the residual sequence $\eta(t)$ is generated by a stationary process, we may set $\sigma_\varepsilon^2\Sigma = D(\eta)$. Then $\Delta'\Sigma\Delta = \Omega_R$, and the solution becomes

$$(50) \qquad x = y - \lambda\Sigma\Delta g.$$

12

Notice that there is no need to find the value of z explicitly, since the value of x can be expressed more directly in terms of $g = \Omega_T^{-1} z$, which is obtained by solving equation (40).

If the residual sequence $\eta(t)$ is nonstationary, then its dispersion matrix is, of course, undefined; and we must find an alternative value for Σ. The problem with this dispersion matrix is attributable to the seasonal operator $S(L)$ of equation (3) which gives rise to a process which, in theory, is unbounded in amplitude. Since the problem arises out of a theoretical assumption of manifest falsity, we should have no qualms about attributing to Σ whatever value seems reasonable. One choice would be to set $\Sigma = I$.

We should also observe that, if y were a vector of T values of a polynomial of degree $d-1$, taken at equally-spaced intervals, then we should have $g = 0$ and therefore $x = y$. That is to say, a polynomial time trend of degree less that d is unaffected by the filter; and this is the most appropriate outcome. If we were to handle the finite-sample problem by any other method, then this result would not be forthcoming, albeit that we might expect it to hold approximately.

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

$$(51) \qquad L(x) = (y - x)'\Sigma^{-1}(y - x) + \lambda x'\Delta\Omega_T^{-1}\Delta'x,$$

where $\lambda = \sigma_\varepsilon^2/\sigma_\nu^2$ as before. We may describe this as the penalised least-squares criterion function. After minimising this function with respect to x, we may use the identity $\Delta'x = z$, which comes from equation (47), and the identity $\Omega_T^{-1}z = g$, which comes from equation (41). Then it will be found that criterion function is minimised by the value specified in (50).

The criterion becomes intelligible when we allude to the assumptions that $y \sim N(\xi, \sigma_\varepsilon^2\Sigma)$ and that $\Delta'\xi = \zeta \sim N(0, \sigma_\nu^2\Omega_T)$; for then it plainly resembles a combination of two independent chi-square variates. The first term of the criterion concerns the goodness of fit of the interpolated trend to the data. The second term imposes a penalty for any roughness in the trend.

This kind of composite criterion function is familiar from the case of the Reinsch smoothing spline [15]. (See de Boor [4], as well) In that context, λ becomes the smoothing parameter which can be adjusted in pursuit of an appropriate trade-off between smoothness and goodness of fit. In the present context, λ, which is specified in (21), is the ratio of the variances of two mutually independent white-noise processes, of which one drives the residue process and the other drives the trend.

5. Trend Extraction via Canonical Decompositions

In an influential article, Hillmer and Tiao [9] have given a complete account of an ARIMA-model-based approach to seasonal adjustment. Their article

proposes a procedure for decomposing a time series into additive components which represent the trend, the seasonal fluctuations, and an irregular noise component. Maravall and Pierce [12] have also analysed these procedures.

In principle, this methodology can be applied to any properly specified seasonal ARIMA model which fits the data. In practice, however, they have provided detailed algebraic decompositions for three models which have been used often to model seasonal economic data. Amongst these is the well-know airline-passenger model of Box and Jenkins [2] which is specified by the equation

$$(52) \qquad (1 - L)(1 - L^s)y(t) = (1 - \theta_1 L)(1 - \theta_2 L^s)\varepsilon(t).$$

In its original application, the model was fitted to the logarithms of a series of monthly observations; and it is usual to take logarithms whenever the trend and the amplitude of the fluctuations surrounding it are growing in an exponential manner.

It is straightforward to explain the role played by the various autoregressive and moving-average factors of this model. The autoregressive factors are subject to the identity $(1 - L)(1 - L^s) = (1 - L)^2 S(L)$, where $S(L) = 1 + L + L^2 + \cdots + L^{s-1}$ is the so-called seasonal sum. Here the factor $(I - L)^2$ pertains to a second-order random walk. Its effect, within an equation of the form $(1 - L)^2 y(t) = \varepsilon(t)$, would be to generate a trend, for which the optimal forecast is a linear extrapolation based only on the two most recent observations.

The factor $S(L)$, which has the roots $\lambda_j = \exp\{i2\pi j/s\}; j = 1, \ldots, s - 1$, is responsible for the pseudo-cyclical seasonal behaviour of the output generated by equation (52). Its effect, within an equation of the form $S(L)y(t) = \varepsilon(t)$, would be to generate a rough cycle which, in the long run, is bounded neither in amplitude nor in phase. The forecast function associated with such a model would be a regular cycle synthesised from $s/2$ sinusoids whose amplitudes and phase angles are determined by the $s - 1$ observations from the most recently observed seasonal cycle together with a zero-mean condition.

The factors $1 - \theta_1 L$ and $1 - \theta_2 L^s$ in the moving-average part of equation (52) serve to counteract some of the effects of the autoregressive unit-root factors $\nabla(L) = I - L$ and $\nabla_s(L) = I - L^s$. The seasonal moving-average operator is subject to the factorisation

$$(53) \qquad \begin{aligned} 1 - \theta_2 L^s &= (1 - \theta_2^{1/s} L)S(\theta_2^{1/s} L) \\ &= (1 - \theta_2^{1/s} L)(1 + \theta_2^{1/s} L + \cdots + \theta_2^{\{s-1\}/s} L^{s-1}). \end{aligned}$$

The leading term of this factorisation joins with the term $1 - \theta_1 L$ in counteracting the random-walk operator $(1 - L)^2$. Their effect is to diminish the power of the random walk at nonzero frequencies, thereby creating a smoother trend. The factor $1 - \theta_2 L^s$, taken as a whole, has the effect of counteracting the
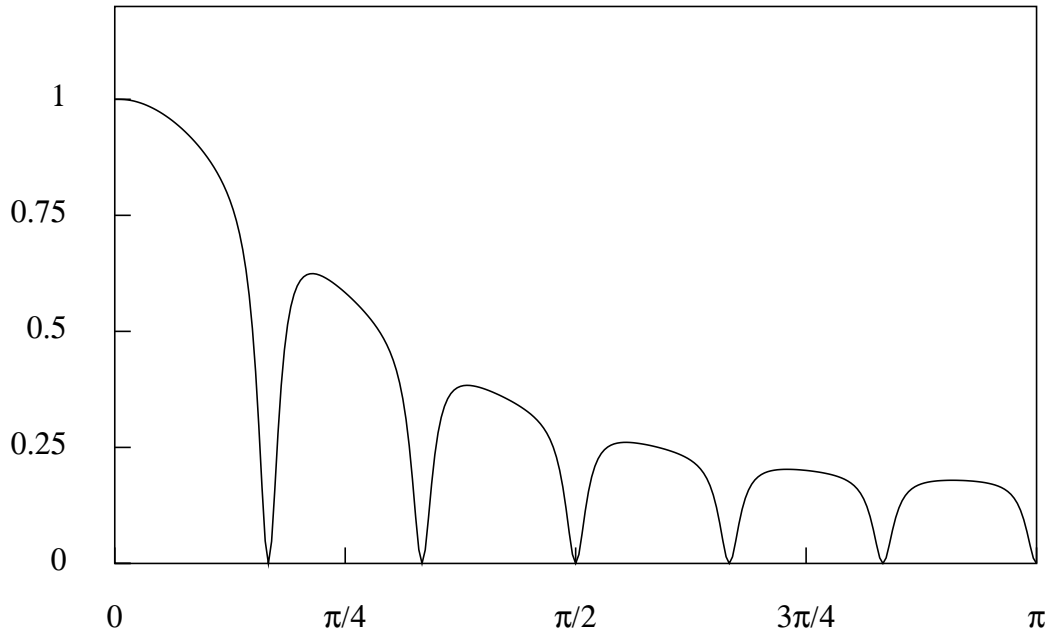
Figure 1. The gain function of the trend-extraction filter based on a partial-fraction decomposition of the airline passenger model.
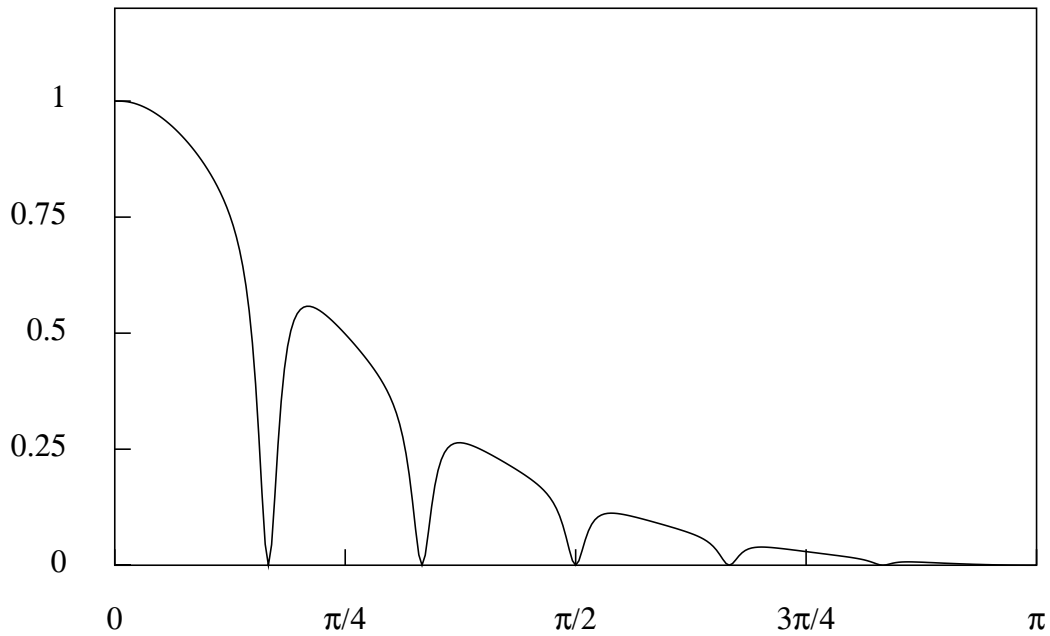
Figure 2. The gain function of the canonical trend-extraction filter based on the airline passenger model.

15

power of the seasonal process except in the vicinities of the seasonal frequencies $w = 2\pi j/s; j = 1, \ldots, s/2$. Thus, as $\theta_2 \to 1$, the widths of the spectral spikes, which are located at these frequencies, are diminished, which both regularises the seasonal cycles which are generated by the model and reduces their phase drift.

Our analysis suggests an alternative way of deploying the parameters θ_1 and θ_2. For if the equation (52) were replaced by the equivalent equation

$$(54) \qquad (1 - L)^2 S(L)y(t) = (1 - \theta_1 L)^2 S(\theta_2^{1/s} L)\varepsilon(t),$$

then the effects of varying θ_1 would be confined to the trend component and the effects of varying θ_2 would be confined to the seasonal component.

The model-based procedure for isolating the components of a data series has three stages. The first stage is to find a partial-fraction decomposition of the autocovariance generating function of the seasonal ARMA model which has been fitted to the data. This takes the form of

$$(55)$$
$$\frac{(1 - \theta_1 z)(1 - \theta_2 z^s)(1 - \theta_2 z^{-s})(1 - \theta_1 z^{-1})}{(1 - z)(1 - z^s)(1 - z^{-s})(1 - z^{-1})}$$
$$= \frac{Q_T(z)}{(1 - z)^2(1 - z^{-1})^2} + \frac{Q_S(z)}{S(z)S(z^{-1})} + \theta_1 \theta_2.$$

When $z = e^{i\omega}$, the term on the LHS of this equation represents the spectral density function of the airline passenger model, whilst the terms of the RHS represent the spectral density functions of its various components. The third term on the RHS represents the uniform spectrum of a white-noise component with a variance of $2\pi\theta_1\theta_2$. It is obtained as a quotient when the numerator of the LHS is divided by the denominator to obtain a proper rational function which is decomposed into the remaining terms of the RHS.

The principle of canonical decomposition is that the variance of the white-noise component should be maximised by assigning to it any white-noise elements which are present in the other two partial-fraction components. This operation of reassignment represents the second stage of the procedure. The outcome is that the revised seasonal and trend components will acquire spectral density functions which are zero-valued somewhere in the frequency range $[0, \pi]$. In particular, the trend spectrum will attain the value of zero at the so-called Nyquist frequency of π.

Detailed algebraic expressions for the elements of equation (55) and for the elements of the revised canonical decomposition have been provided by Hillmer and Tiao [9]. Such derivations call for care and stamina; and it is easier to rely upon a computational approach for finding the coefficients of the partial fractions, which is essentially a matter of solving a set of linear equations.

The third and final stage of the procedure for isolating the components of the data is to form the appropriate filters and to apply them to the data series. The recommended techniques for applying such filters have been discussed at length already in Section 4 of this paper. Here we shall do no more than present the gain function of the model-based filter for extracting the trend component. This is given in Figure 2.

The profile of the gain of the filter contains a sequence of notches which are at the seasonal frequency of $\omega = \pi/6$ and at the harmonic frequencies of $\pi j/6; j = 2, 3, \ldots, 6$. These notches, which serve to exclude the seasonal component from the trend, are the effects of the zeros of the filter.

Apart from the notches, the profile shows a gradual transition from the value of unity at the frequency $\omega = 0$ to the value of zero at the Nyquist frequency of $\omega = \pi$. Thus the estimated trend comprises a wide range of frequencies. This feature is at variance with a common definition of a trend which proposes that it should contain only a limited set of low-frequency components— with the maximum frequency falling short of the seasonal frequency of $\omega = \pi/6$.

It is clear from Figure 1 that an estimated trend which is based only on the partial fraction decomposition of the seasonal ARMA model, and which pays no heed to the principal of canonical decomposition, is liable to embody a significant proportion of high-frequency noise.

6. Trend Estimation via Structural Time-Series Models

The structural time-series model, which has been proposed by Harvey and Todd [8], can be written in the form of

$$(56) \qquad y(t) = \frac{\zeta(t-1)}{\nabla^2(L)} + \frac{\eta(t)}{\nabla(L)} + \frac{\omega(t)}{S(L)} + \varepsilon(t),$$

where $\zeta(t)$, $\eta(t)$, $\omega(t)$ and $\varepsilon(t)$ are mutually independent white-noise processes. By combining the leading terms on the RHS, the equation may be rewritten as
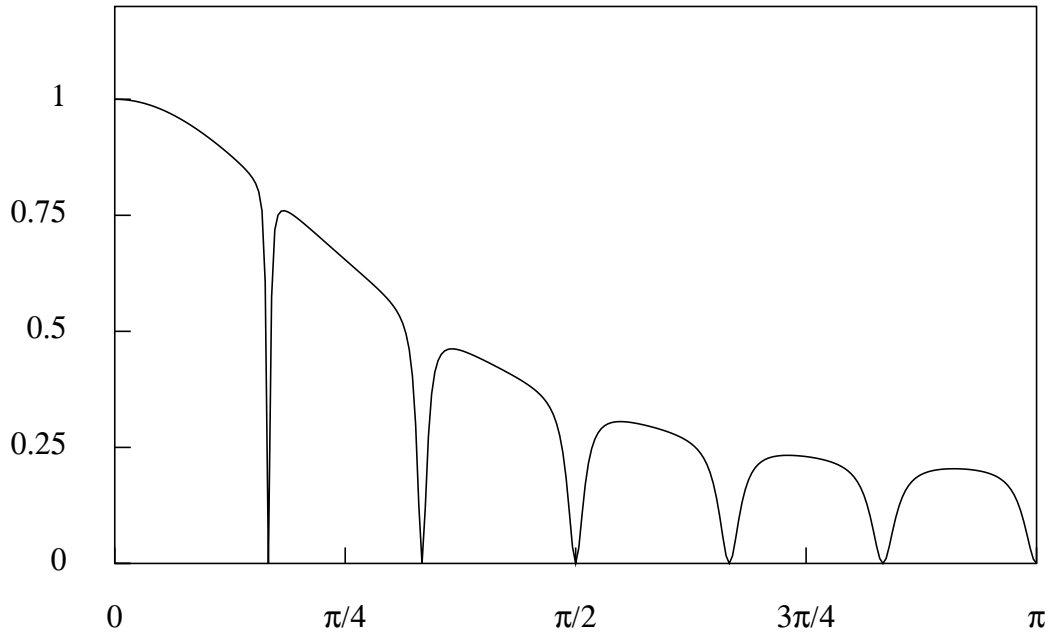
$$(57) \qquad y(t) = \frac{\xi(t)}{\nabla^2(L)} + \frac{\omega(t)}{S(L)} + \varepsilon(t),$$

where

$$(58) \qquad \xi(t) = \nabla(L)\eta(t) + \zeta(t-1) = (1 - \mu L)\nu(t)$$

follows a first-order moving-average process.

The first term on the RHS of (57) represents the trend and the second term represents the seasonal fluctuations. It can be seen that the trend follows an integrated moving-average IMA(2, 1) model which is a second-order random

17

Figure 3. The gain function of the trend-extraction filter based on
a structural time-series model fitted to the airline passenger data.
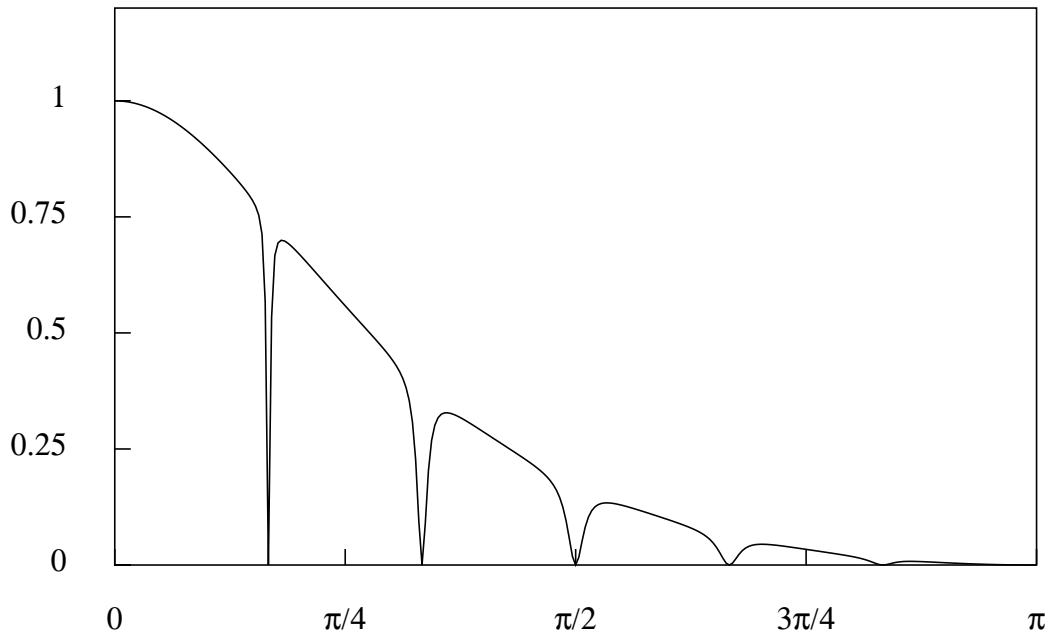
Figure 4. The gain function of a canonical version of the
trend-estimation filter based on a structural times-series model.

walk driven by a first-order moving-average forcing function. A similar trend process is implicit in the airline passenger model described by equation (52).

The attraction of the structural model is that it is already decomposed into the appropriate components. The decomposition is not a canonical one, but there is no reason why white-noise elements should not be subtracted from the trend and the seasonal components and reassigned to the irregular component.

Figure 3 displays the gain of the trend-extraction filter based on a structural model which has been estimated from the airline-passenger data of Box and Jenkins [2]. It is clear that the filter will include in the estimated trend a substantial amount of high-frequency noise which would be excluded by the canonical model-based filter represented in Figure 2. In consequence, the estimated trend will have a very rough appearance.

Figure 4 displays the gain of an amended filter derived by eliminating the white-noise element from the model of the trend. The profile of the amended filter is similar to that of the canonical filter displayed in Figure 2.

## 7. Trend Extraction via Square-Wave Filters

The third approach to trend estimation which we shall consider is commonly described as the model-free approach. Since many of the filters in question can be derived from a model of a stochastic process, it is perhaps misleading to describe them as model-free. Nevertheless, such models are usually regarded only as heuristic devices; and they do not always purport realistically to represent the sequences which are to be filtered.

In an heuristic model of the sort which is used in deriving an estimate of the trend, we are liable to find only stylised representations of the other components. For, if the frequency range which defines the trend is segregated from frequency ranges of the remaining components, and if the intention is to suppress these components, then it should be unnecessary to represent them with much realism.

The question of whether or not the trend is segregated from the other components depends partly on how we choose to define it and partly on the nature of the series itself. Figure 5 shows the logarithms of the airline passenger data together with an interpolated linear trend. Figure 6 shows the periodogram of the data, whilst Figure 7 shows the periodogram of the interpolated linear trend.

There may be some surprise at the fact that the periodogram of the linear trend is not wholly confined to the zero frequency. Its form is explained once it is recognised that the underlying Fourier synthesis, whose coefficients are incorporated in the periodogram, is an approximation to a periodic sawtooth function, of which the linear function defined on the interval $[0, T)$ is but one segment.
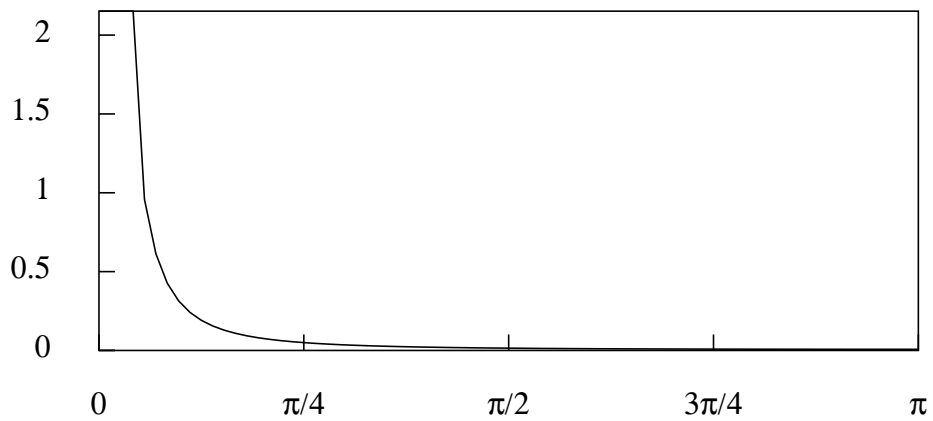
As we have shown in Section 4, a linear trend will be preserved by any
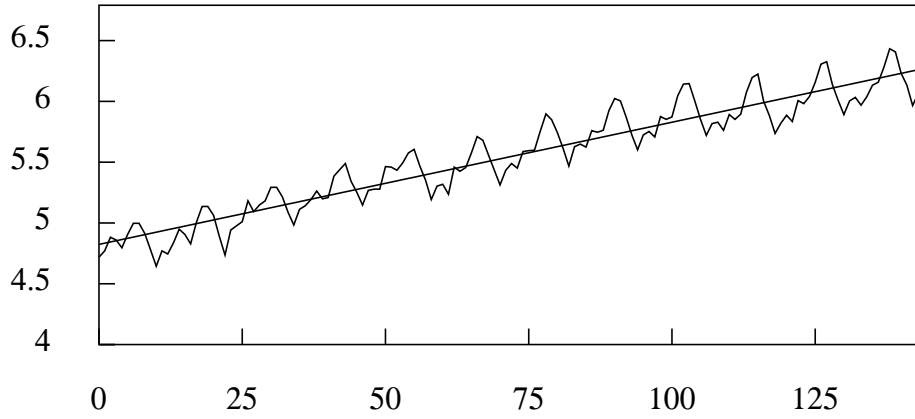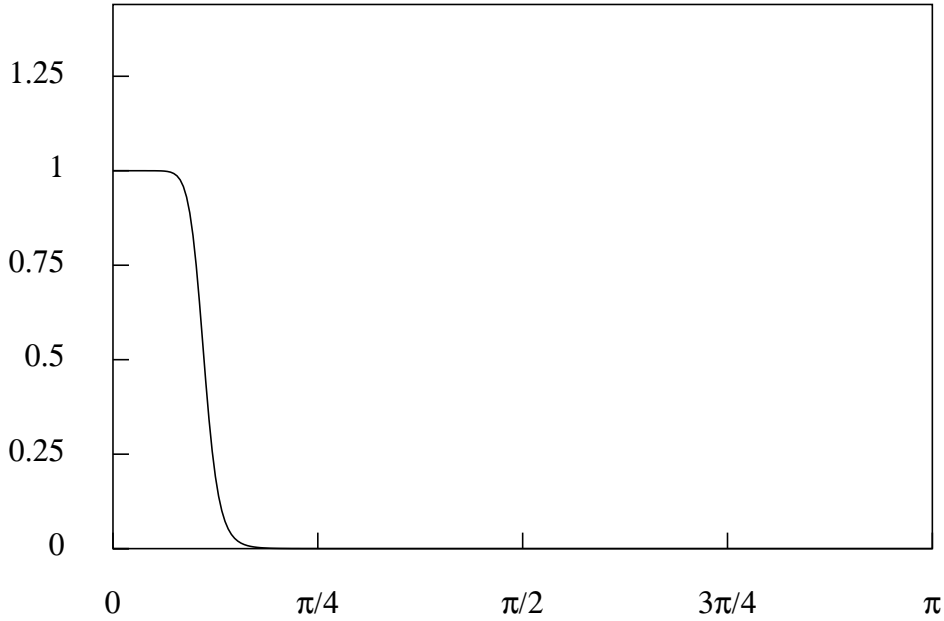
Figure 5. The logarithms of 144 monthly observations on the number
of international airline passengers with an interpolated linear trend.
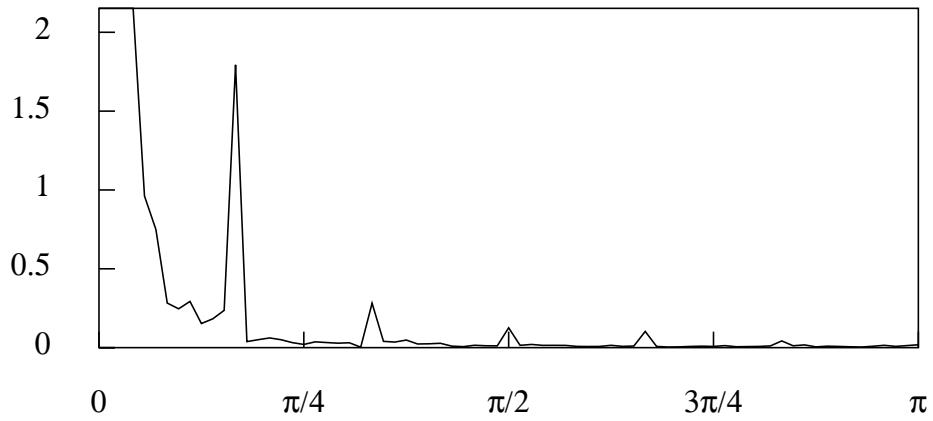
Figure 6. The periodogram of the airline passenger data.

Figure 7. The periodogram of the linear trend which
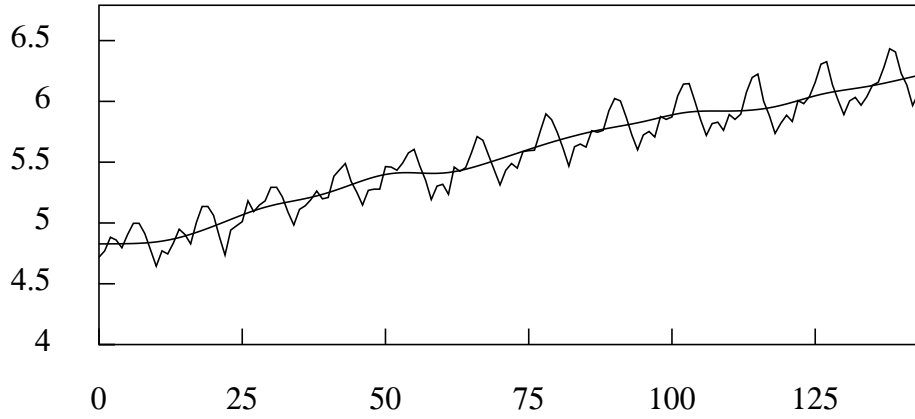has been fitted to the airline passenger data.

Figure 8. The gain of the 6th order Butterworth lowpass
filter with a nominal cut-off frequency of $\omega_c = \pi/9$ degrees.

of the finite-sample filters which embodies the assumption that $d = 2$ in the
equation (2) which represents the trend component. There is evidence that
the trend component of the airline passenger data contains some additional
elements whose frequencies lie in the interval between zero and the seasonal
frequency of $\omega = \pi/6$. Indeed, this is evident in the periodogram of the residual
sequence obtained by fitting the linear function.

We are therefore motivated to find a linear filter which will estimate the
trend by preserving those elements, which are additional to the linear trend,
whose frequencies are bounded by a value ω_c which is slightly below the seasonal
frequency. At the same time, the filter should suppress all elements which
are not part of the linear trend whose frequencies exceed this value. Such
a filter should be effective not only in the present circumstances but also in
circumstances where it is quite inappropriate to model the trend by fitting an
analytic function.

One filter which serves the purpose of isolating a well-defined range of
frequencies is the so-called Butterworth square-wave filter (See, for example,
Pollock [14]). The filter can be derived from an heuristic model represented by
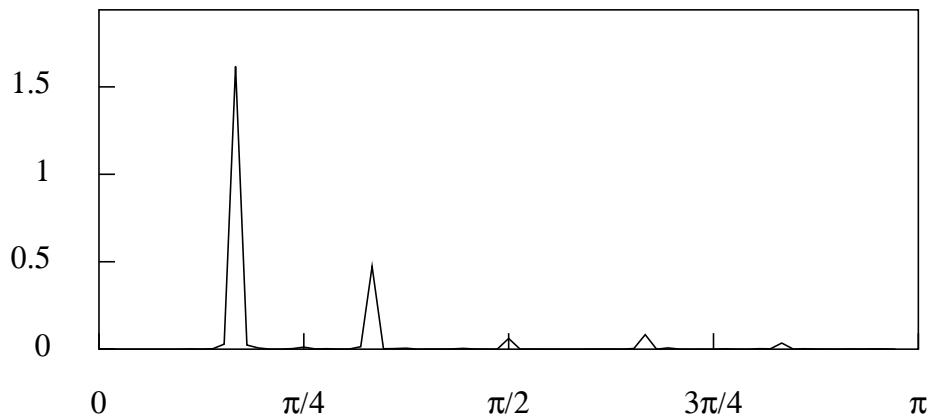the equation

$$(59) \qquad \begin{aligned} y(t) &= \xi(t) + \eta(t) \\ &= \frac{(1+L)^n}{(1-L)^d}\nu(t) + (1-L)^{n-d}\varepsilon(t). \end{aligned}$$

21

Figure 9. The airline passenger data with an interpolated trend estimated by a Butterworth filter with $n = 6$ and $\omega_c = \pi/9$.
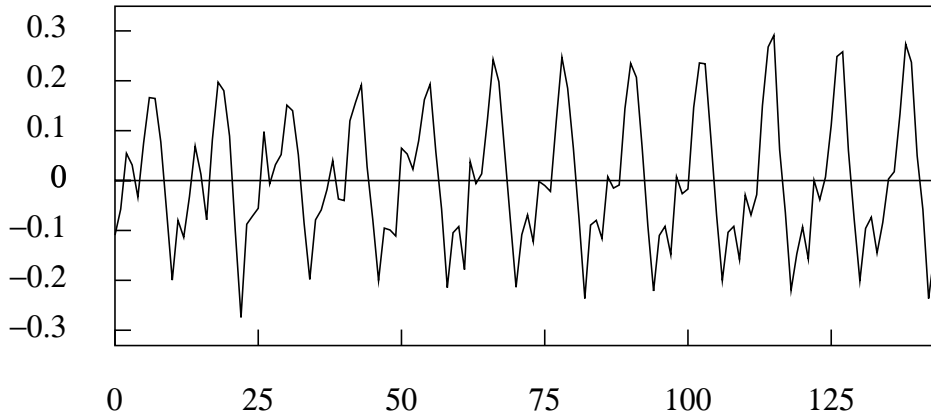
Figure 10. The residual sequence obtained by de-trending the airline passenger data with a Butterworth filter.

Figure 11. The periodogram of the residuals obtained by de-trending the airline passenger data with a Butterworth filter.

22

The Wiener–Kolmogorov form of the resulting trend-extraction filter is

$$
(60) \qquad \psi_T(L) = \frac{(1+L)^n(1+L^{-1})^n}{(1+L)^n(1+L^{-1})^n + \lambda(1-L)^n(1-L^{-1})^n},
$$

where $\lambda = \sigma_\varepsilon^2/\sigma_\nu^2 = \{1/\tan(\omega_c)\}^{2n}$.

Figure 8 shows the gain of the filter, whilst Figures 9–11 show the effects of applying the finite-sample version of the filter to the airline passengers data. In this case, the parameters of the filter are $d = 2, n = 6$ and $\omega_c = \pi/9$. It is evident from Figure 8, which represents the gain of the filter, that the trend which is seen in Figure 9 is composed of a set of elements which fall within a strictly limited frequency range. This definition of a trend contrasts markedly with the definitions which are implicit in the two model-based approaches to trend estimation where the trend is composed of a broad range of frequencies excluding only the seasonal frequency and its harmonics.

8. Summary and Conclusions

In this paper, we have endeavoured to provide an account of some of the principal methods of econometric trend estimation which are based upon statistical models of the processes generating the data. We have shown that there is a single mathematical framework which can accommodate quite disparate approaches to the problem.

Trend estimation is often regarded as a difficult task which is beset by technical complexity. The complications have two sources. In the first place, there are the difficulties of the structural ARMA model-based approach of Hillmer and Tiao [9] which depends upon the partial-fraction decomposition of a reduced-form seasonal ARIMA model. Matters are greatly simplified when the alternative structural approach of Harvey and Todd [8] is pursued which represents the structural components explicitly and which avoids the need to recover them from an ARIMA model. The difficulties disappear altogether if one adopts an heuristic model, such as the model which underlies the square-wave filter, which contains only a simplified and stylised representation of the trend-free components.

The second source of difficulty concerns the need to adapt the classical Wiener–Kolmogorov theory of signal extraction so that it can be applied to data series which are both heavily trended and of strictly limited duration. Recently, a number of solutions to the finite-sample problem have been proposed within the context of the Kalman filter. In this paper, we are proposing a simple solution which appears to be definitive. The solution can be accommodated within the framework of the Kalman filter, but it also has some other, quite separate, antecedents.

We have taken a liberal approach to the matter of how a trend is best defined. Our own prescription, that it should comprise a set of elements falling

within a limited range of frequencies, is clearly at variance with the definitions which are implicit in the two approaches which we have examined which are based on structural models. If a de-trended series is to be used as an explanatory variable in a further analysis, then there may be some advantages in the method of de-trending which has least effect upon the trend-free components. Such considerations would favour the square-wave de-trending filter.

9. References

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

[2] Box, G.E.P., and G.M. Jenkins, (1976), *Time Series Analysis: Forecasting and Control, Revised Edition,* Holden Day, San Francisco.

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

[4] de Boor, C., (1978), *A Practical Guide to Splines,* Springer Verlag, New York.

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

[6] Gómez, V. and A. Maravall, (1994), Estimation Prediction and Interpolation for Nonstationary Series with the Kalman Filter, *Journal of the American Statistical Association,* **89,** 611–624.

[7] Harvey, A.C., (1989), *Forecasting, Structural Time Series Models and the Kalman Filter,* Cambridge University Press, Cambridge.

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

[11] 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.

[12] Maravall, A., and D.A. Pierce, (1987), A Prototypical Seasonal Adjustment Model, *Journal of Time Series Analysis,* **8,** 177–193.

[13] Pierce, D.A., (1979), Signal Extraction in Nonstationary Time Series, *The Annals of Statistics,* **6,** 1303–1320.

[14] Pollock, D.S.G., (1997), *Data Transformations and De-Trending in Econometrics, in Heij, C. et al., Systems Dynamics in Economic and Financial Models,* John Wiley and Sons.

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

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

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