FILTERS FOR SHORT NONSTATIONARY SEQUENCES

by D.S.G. Pollock

GREQAM: Groupement de Recherche en
Economie Quantitative d'Aix–Marseille

and

Queen Mary and Westfield College
University of London

Email: stephen_pollock@sigmapi.u-net.com

This paper describes a methodology for implementing bidirectional frequency-selective filters in cases where the data sequence is short and nonstationary. A simple method is proposed for dealing with the start-up problem. The method has a firm theoretical basis and it is computationally efficient.

Key words: Signal extraction, Linear filtering, Frequency-domain analysis, Trend estimation;

JEL Classification: C22

1. Introduction

In the classical Wiener–Kolmogorov theory of signal extraction (see [16] and [12]), it is envisaged that the data are generated by a stationary stochastic process and that they form a lengthy sequence. The data process is often depicted as a combination of a signal process, with a low-frequency spectrum, and a noise process, with a high-frequency spectrum. Sometimes, it envisaged that the noise process has a uniform spectrum. This is appropriate whenever the signal process is obscured by a sequence of independently and identically distributed errors of observation. In that case, the noise is aptly described as a nuisance component.

The Wiener–Kolmogorov filters are often applied in circumstances which are at odds with the classical assumptions. The data sequence may be a short one with barely more than 50 observations. It might also show a underlying trend which describes an irregular or a highly nonstationary trajectory.

The purposes for which the filters are used are sometimes the opposite of those envisaged by the classical theory. If the interest centres on the high-frequency component, then it is the low-frequency trend which might be regarded as the nuisance component. Often, the trend and the residue are of equal interest.

The purpose of this paper is to describe the manner in which the Wiener–Kolmogorov theory can be adapted to cope with short, trended sequences. The resulting methodology can be used both for the purpose of removing the trend from a data sequence and for the purpose of revealing the trend more clearly. Two areas of application which spring to mind are meteorology and economics.

In meteorology, there is currently much interest in the problems of discerning the underlying trends in various global climatic indices such as the index of the annual temperature anomalies in the global hemispheres (see Figure 1), or the index of atmospheric carbon-dioxide concentration (see Figure 2). In the case of the temperature anomalies, it is a matter of stripping away the noise of the short-term effects and of the measurement errors so as to clarify the trend. In the case of the carbon dioxide index, we see a remorselessly upwards trend which is surrounded by annual fluctuations of a remarkable regularity. Here, both the trend and the annual cycles which surround it are of interest.

Seasonal effects which are analogous to those present in the carbon dioxide index can be found in many indices of economic activity (see Figure 3). Whether the seasonal fluctuations are of interest in their own right or whether they are merely a nuisance will depend upon the perspectives of the individual economist. A macro economist might be interested, primarily, in the long-term trend of an index of unemployment. A labour economist might be interested in the patterns of hiring and firing which would be revealed by de-trending the series and by isolating its seasonal fluctuations.

2. Filtering Nonstationary Sequences

The task of adapting classical signal-extraction filters to limited samples from nonstationary processes has caused difficulties and perplexity. Problems often arise from not knowing how to supply the initial conditions with which to start a recursive filtering process. By choosing inappropriate starting values, one can generate so-called transient effects which are liable, in fact, to affect all of the processed values.

One common approach to the problem of the start-up conditions relies upon the ability to extend the sample by forecasting and backcasting. The additional extra-sample values can be used in a run-up to the filtering process wherein the filter is stabilised by providing it with a plausible history, if it is working in the direction of time, of with a plausible future, if it is working in reversed time. Sometimes, very lengthy extrapolations are called for (see Burman [6], for example).

An alternative approach to the start-up problem is to estimate the requisite initial conditions. Some of the methods which follow this approach have been devised within the context of the Kalman filter and the associated smoothing algorithms—see Ansley and Kohn [1] and de Jong [7], for example.

The approach which we shall adopt in this paper avoids the start-up problem by applying the filter, in the first instance, to a version of the data sequence which has been reduced to stationarity by repeated differencing.

We can proceed to find an estimate of the residual sequence by cumulating its differenced version. Here, if the residual sequence has a significant degree of serial dependence, we can profit from some carefully estimated start-up values to set the process of cumulation in motion; but, in fact, as we shall see, one can avoid finding such values explicitly. However, if the residual sequence is

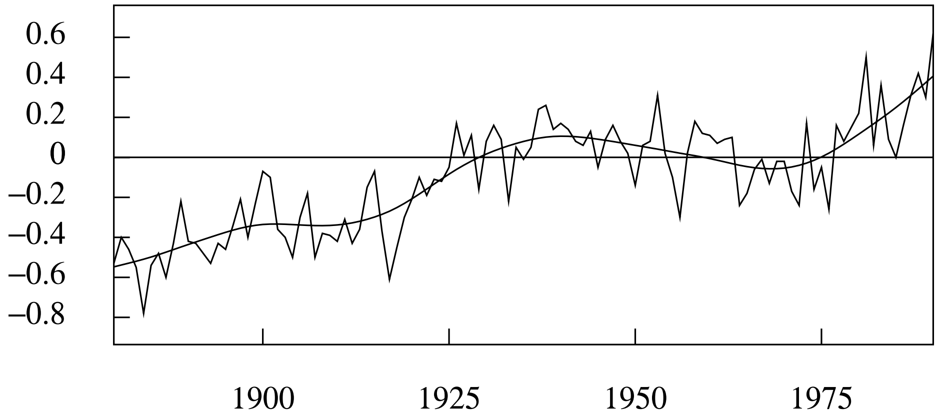

Figure 1. Northern hemisphere temperature anomalies 1880–1990.

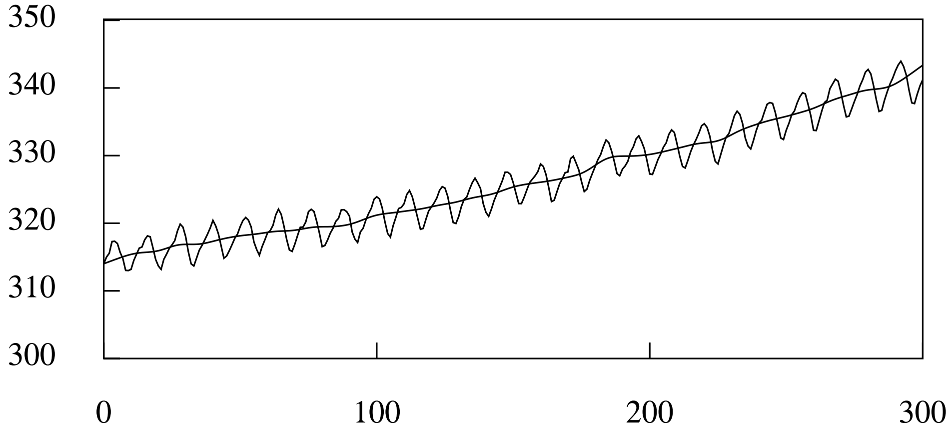

Figure 2. Monthly atmospheric carbon dioxide concentrations in parts per million by volume from Jan. 1958 to Jan. 1984.

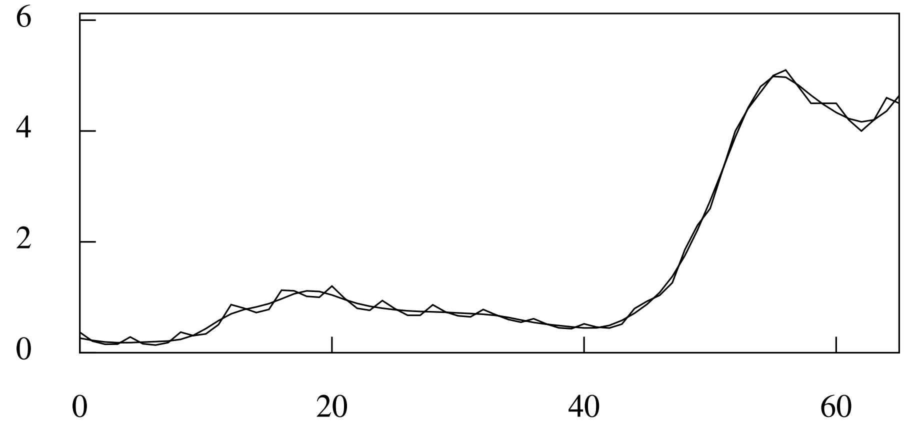

Figure 3. The quarterly figures on Swiss unemployment from 1980.1 to 1996.2.

generated by a white-noise process, or if it has only a weak serial dependence, then it may be acceptable to replace these start-up values by their unconditional expectations which are zeros.

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

To clarify these matters, we may begin by considering a general model of the processes which have generated the data. We shall assume that the trend or signal sequence $\xi(t)$ is generated by a nonstationary autoregressive integrated moving-average (ARIMA) process and that the residual component $\eta(t)$, which is its complement, is generated by an ordinary stationary autoregressive moving-average (ARMA) process. Thus the data sequence $y(t)$, 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

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

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

Here $\psi_T(L)$ and $\psi_R(L)$ are proper rational functions of the lag operator L which have all of their poles and zeros lying outside the unit circle. It is assumed that they have no factors in common. The lag operator has the effect that $Lx(t) = x(t-1)$. The operator $(1-L)^{-d}$ —which is apt to be described as a summation or integration operator—is wholly responsible for the nonstationary character of $\xi(t)$.

Equation (1) may be multiplied throughout by the product of the denominators to give

$$\begin{aligned} g(t) &= \delta_T(L)\nu(t) + \delta_R(L)\varepsilon(t) \\ &= \zeta(t) + \kappa(t), \end{aligned} \tag{2}$$

where $\delta_T(L)$ and $\delta_R(L)$ polynomials of finite degree. In the sequel, we shall assume that both $\psi_T(L)$, $\psi_R(L)$ are finite-degree polynomials. This assumption entails only a small loss of generality and it serves to simplify the exposition. It follows that

$$\begin{aligned} &g(t) = (I-L)^d y(t), \\ &\zeta(t) = \delta_T(L)\nu(t) = \psi_T(L)\nu(t), \\ &\kappa(t) = \delta_R(L)\varepsilon(t) = (1-L)^d\psi_R(L)\varepsilon(t). \end{aligned} \tag{3}$$

In the more general case, where $\psi_T(L)$ and $\psi_R(L)$ are rational functions, there would be an additional factor in the mapping from $y(t)$ to $g(t)$ compounded from the denominators of $\psi_T(L)$ and $\psi_R(L)$.

The Wiener–Kolmogorov theory of statistical signal extraction, as expounded by Whittle [17] for example, indicates that an estimate $z(t)$ of the sequence $\zeta(t)$ can be obtained from the stationary sequence $g(t)$ via a filtering operation which is described by the equation

$$\begin{aligned} & z(t) = \beta_T(L)g(t) = \frac{\delta_T(F)\delta_T(L)}{\gamma(F)\gamma(L)}g(t), \\ (4) \qquad & \text{where} \quad \gamma(F)\gamma(L) = \delta_T(F)\delta_T(L) + \lambda\delta_R(F)\delta_R(L), \\ & \text{with} \quad \lambda = \frac{\sigma_\varepsilon^2}{\sigma_\nu^2}. \end{aligned}$$

Here $F = L^{-1}$ is the forward shift operator which is the inverse of the lag operator and which has the effect that $Fp(t) = p(t+1)$.

The filter is a bidirectional one which must necessarily be applied in at least two passes running forward and backward through the data sequence. Thus the sequence $z(t)$ can be generated via the following recursive processes:

$$(5) \quad \text{(i)} \quad \gamma(L)f(t) = \delta_T(L)g(t) \qquad \text{and} \qquad \text{(ii)} \quad \gamma(F)z(t) = \delta_T(F)f(t).$$

Process (i) generates an intermediate output $f(t)$ from the forwards pass of the filter, whereas process (ii) generates the final output $z(t)$ in a backwards pass.

An estimate $k(t)$ of the sequence $\kappa(t)$ can be obtained from $g(t)$ via similar operations which are summarised by the equation

$$(6) \qquad k(t) = \beta_R(L)g(t) = \lambda\frac{\delta_R(F)\delta_R(L)}{\gamma(F)\gamma(L)}g(t).$$

This can also be realised in the manner of equation (5). However, several alternative implementations are also available. Consider

$$(7) \quad \begin{aligned} &\text{(i)} \quad \gamma(L)p(t) = g(t), && \text{(ii)} \quad \gamma(F)b(t) = p(t), \\ &\text{(iii)} \quad z(t) = \delta_T(F)\delta_T(L)b(t), && \text{(iv)} \quad k(t) = \lambda\delta_R(F)\delta_R(L)b(t). \end{aligned}$$

Here steps (i) and (ii) represent forwards and backward recursive processes which are common to the estimation of $\zeta(t)$ and $\kappa(t)$. Steps (iii) and (iv), each of which can realised in a single forward or backwards pass, generate the estimates of $\zeta(t)$ and $\kappa(t)$ respectively.

The lowpass filter of equation (4) will generate the minimum mean-square-error estimate of the stationary sequence $\zeta(t)$ provided that the smoothing parameter has the value of $\lambda = \sigma_\varepsilon^2/\sigma_\nu^2$. However, as Bell [2] has established, the Wiener–Kolmogorov theory applies equally to nonstationary processes. Thus, if it were applied to the nonstationary series $y(t)$, the lowpass filter would generate the minimum mean-square-error estimate of the nonstationary trend sequence $\xi(t)$.

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

$$\begin{aligned}\xi(t) &= y(t) - \eta(t)\\ &= y(t) - \frac{\kappa(t)}{(1-L)^d}.\end{aligned} \tag{8}$$

The equation suggests that we may begin by estimating the stationary sequence $\kappa(t)$ by applying the filter of (6) to $g(t)$ which is the differenced version of the data sequence. Thereafter, an estimate of the stationary sequence $\eta(t)$ can be obtained by a d-fold process of cumulation. Finally, the estimate of $\xi(t)$ can be obtained by a simple subtraction.

3. Filtering Short Sequences

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

$$\begin{aligned}y &= \xi + \eta\\ &= x + h,\end{aligned} \tag{9}$$

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

$$E(\eta) = 0 \qquad \text{and} \qquad D(\eta) = \sigma_\varepsilon^2\Sigma. \tag{10}$$

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

To find the finite-sample the counterpart of equation (2), we need to represent the d-th difference operator $(1-L)^d = 1 + \delta_1 L + \cdots + \delta_d L^d$ in the form of a matrix. Therefore, let the identity matrix of order T be denoted by

$$I_T = [e_0, e_1, \ldots, e_{T-1}], \tag{11}$$

where e_j represents a column vector with a unit in the position j and with zeros elsewhere. Then the finite-sample lag operator is the matrix

$$L_T = [e_1, \ldots, e_{T-1}, 0], \tag{12}$$

which has units on the first subdiagonal and zeros elsewhere. This matrix is formed by deleting the leading vector of the identity matrix and by appending a zero vector to the end of the array. The lag-operator polynomials which have been used in the foregoing analysis can be converted to matrix operators of order T simply by replacing the lag operator L by the matrix L_T. Thus, the matrix which takes the d-th difference of a vector of order T is given by $\Delta = (I - L_T)^d$.

Taking differences within a vector entails a loss of information. Thus, if $\Delta = [Q_*, Q]'$, where Q'_* has d rows, then the d-th differences of the vector $y = [y_0, \ldots, y_{T-1}]'$ are the elements of the vector $g = [g_d, \ldots, g_{T-1}]'$ which is found in the equation

$$\begin{bmatrix} g_* \\ g \end{bmatrix} = \begin{bmatrix} Q'_* \\ Q' \end{bmatrix} y. \tag{13}$$

The vector $g_* = \Delta'_* y_*$ in this equation, which is a transform of the vector $y_* = [y_0, \ldots, y_{d-1}]'$ of the initial elements, is liable to be discarded. The matrix of the transformation is the operator $\Delta_* = (I - L_d)^d$; and it follows that $Q'_* = [\Delta_*, 0]$.

Premultiplying equation (9) by Q' gives

$$\begin{aligned} g = Q'y &= \zeta + \kappa \\ &= z + k, \end{aligned} \tag{14}$$

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

$$E(\zeta) = 0 \qquad \text{and} \qquad D(\zeta) = \sigma_\nu^2 \Omega_T, \tag{15}$$

and those of κ by

$$\begin{aligned} E(\kappa) = 0 \qquad \text{and} \qquad D(\kappa) &= Q'D(\eta)Q \\ &= \sigma_\varepsilon^2 Q'\Sigma Q = \sigma_\varepsilon^2 \Omega_R, \end{aligned} \tag{16}$$

where both 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 matrices are, respectively, $\delta_T(z)\delta_T(z^{-1})$ and $\delta_R(z)\delta_R(z^{-1})$, where δ_T and δ_R are defined in (3).

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

$$\begin{aligned} E(\zeta|g) &= E(\zeta) + C(\zeta, g)D^{-1}(g)\{g - E(g)\} \\ &= \Omega_T(\Omega_T + \lambda\Omega_R)^{-1}g = z. \end{aligned} \tag{17}$$

The optimal predictor k of $\kappa = Q'\eta$ is given, likewise, by

$$\begin{aligned} E(\kappa|g) &= E(\kappa) + C(\kappa, g)D^{-1}(g)\{g - E(g)\} \\ &= \lambda\Omega_R(\Omega_T + \lambda\Omega_R)^{-1}g = k. \end{aligned} \tag{18}$$

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

The estimates are calculated, first, by solving the equation

$$(\Omega_T + \lambda\Omega_R)b = g \tag{19}$$

for the value of b and, thereafter, by finding

$$z = \Omega_T b \qquad \text{and} \qquad k = \lambda\Omega_R b. \tag{20}$$

The solution of equation (19) is found via a Cholesky factorisation which sets $\Omega_T + \lambda\Omega_R = GG'$, where G is a lower-triangular matrix. The system $GG'b = g$ may be cast in the form of $Gp = g$ and solved for p. Then $G'b = p$ can be solved for b.

Observe that the generating function for the matrix GG' is the polynomial $\gamma(z)\gamma(z^{-1})$ defined in (4). Moreover, the solution via the Cholesky factorisation is a finite-sample analogue of a process of bidirectional filtering which finds the sequence $b(t)$ via the recursive processes of (7)(i) and (7)(ii). The equations of (20) correspond, respectively, to the filtering operations of (7)(iii) and (7)(iv).

4. Recovering the Trended Sequence

Our object is to recover from z or from k an estimate x of the trend vector ξ. This would be conceived, ordinarily, as a matter of integrating the vector z d times via a simple recursion which depends upon d initial conditions. However, for reasons of numerical stability, we prefer the alternative approach in which h, which is the estimate of η, is first recovered from $k = Q'h$, which is the estimate of $\kappa = Q'\eta$. Then $x = y - h$ can be found by subtraction.

We shall examine two methods of recovering the trend which take this alternative approach. The first method might be described as the direct method. The second method, which is the more intriguing, is superior in terms of its numerical stability and its ease of implementation. We shall demonstrate the mathematical equivalence of the two methods and, thereafter, we shall show that they may be derived from an appropriate statistical criterion.

Let $M = [S_*, S]$ denote the inverse of the matrix $\Delta = [Q_*, Q]'$ of equation (13). Then the estimate h of the residual vector η is given by the equation

$$h = S_* k_* + Sk, \tag{21}$$

wherein k_* is a vector of initial values, or constants of integration. We seek a value for k_* which will bring the trend into a close alignment with the data. Therefore, the vector is determined by evaluating the following generalised least-squares criterion:

$$\text{Minimise} \quad (S_* k_* + Sk)'\Sigma^{-1}(S_* k_* + Sk) \quad \text{with respect to} \quad k_*. \tag{22}$$

The minimising value is

$$k_* = -(S_*'\Sigma^{-1}S_*)^{-1}S_*'\Sigma^{-1}Sk. \tag{23}$$

On defining the operator $P_* = S_*(S_*'\Sigma^{-1}S_*)^{-1}S_*'\Sigma^{-1}$, which is an idempotent projector, we may write

$$h = (I - P_*)Sk; \tag{24}$$

and the estimated trend is given by

$$\begin{aligned} x &= y - h \\ &= y - (I - P_*)Sk. \end{aligned} \tag{25}$$

Observe that, if $\Sigma = I$, which is to say that the vector η is generated by a white-noise process, and if $d = 1$, which means that $S_*' = [1, \ldots, 1]$ becomes the summation vector, then k_* is just the ordinary average of the elements of Sk. This average is an estimate of a zero-valued expectation; and it would be appropriate to set the starting value to zero. In general, whenever η is from a white-noise process it is acceptable to use zeros for the starting values.

On the other hand, if $\sigma_\varepsilon^2\Sigma = D(\eta)$ is the dispersion matrix of a finite order moving-average process, then the inverse matrix Σ^{-1} has nonzero elements in every location. Given that Σ^{-1} is of order $T \times T$, it is liable to be large. This can create problems in computing the value of k_* via the formula of (23).

We can avoid the difficulties of estimating the value of k_* by seeking the solution to the following problem:

$$\text{Minimise} \quad (y - x)'\Sigma^{-1}(y - x) \quad \text{subject to} \quad Q'x = z. \tag{26}$$

This criterion poses the problem of finding an estimated trend vector which is closely aligned to the data and which has a differenced value equal to the filtered value z generated by equation (20).

The problem is addressed by evaluating the Lagrangean function

$$L(x, \mu) = (y - x)'\Sigma^{-1}(y - x) + 2\mu'(Q'x - z). \tag{27}$$

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

$$\Sigma^{-1}(y - x) - Q\mu = 0. \tag{28}$$

Premultiplying by $Q'\Sigma$ gives

$$Q'(y - x) = Q'\Sigma Q\mu. \tag{29}$$

But, from (19) and (20), it follows that

$$\begin{aligned} Q'(y - x) &= g - z \\ &= \lambda\Omega_R b = \lambda Q'\Sigma Q b, \end{aligned} \tag{30}$$

whence, from (29), we get

$$\begin{aligned} \mu &= (Q'\Sigma Q)^{-1}Q'(y - x) \\ &= \lambda b. \end{aligned} \tag{31}$$

Putting the final expression for μ into (28) gives

$$x = y - \lambda\Sigma Q b. \tag{32}$$

This is our preferred solution to the problem of estimating the trend vector ξ; and the equation indicates the most efficient way of computing the value of x. The advantage of this method of recovering x is that there is no need to invert the matrix Σ. Moreover, the matrix Q typically comprises only a small handful of distinct elements. Therefore the computation is relatively undemanding.

The vector b, upon which the computation of x depends, is obtained as the solution to equation (19). From b is formed the vector $\lambda\Sigma Q b = h$ of equation (32) which represents the estimate of the residue vector. There is no need to find an explicit value for $k = Q'h$ which is the differenced version of the vector.

Observe that, when the expression for λb from (31) is substituted into (32), the result is

$$x = y - \Sigma Q(Q'\Sigma Q)^{-1}(Q'y - z). \tag{33}$$

On defining $P_Q = \Sigma Q(Q'\Sigma Q)^{-1}Q'$, which is an idempotent projector, the equation can be written as

$$\begin{aligned} x &= y - P_Q(y - x) \\ &= y - P_Q h. \end{aligned} \tag{34}$$

This is to be compared with equation (25) which can be written as

$$\begin{aligned} x &= y - (I - P_*)(S_*k + Sk) \\ &= y - (I - P_*)h, \end{aligned} \tag{35}$$

which follows from the fact that $(I - P_*)S_*k_* = 0$ and from the identity of (21). The equivalence of equations (34) and (35) is established by showing that $P_Q = I - P_*$. The appendix of the paper contains a demonstration. Thus, the two approaches which originate in the seemingly distinct criteria of (22) and (26) are, in fact, mathematically equivalent.

It is notable that there is a criterion function which will enable us to derive the equation (32) of the trend estimation filter in a single step. The function is

$$L(x) = (y - x)'\Sigma^{-1}(y - x) + \lambda x'Q\Omega_T^{-1}Q'x, \tag{36}$$

wherein $\lambda = \sigma_\varepsilon^2/\sigma_\nu^2$, as before. This is minimised by the value specified in (32). The criterion function becomes intelligible when we invoke to the assumptions that $y \sim N(\xi, \sigma_\varepsilon^2\Sigma)$ and that $Q'\xi = \zeta \sim N(0, \sigma_\nu^2\Omega_T)$; for then it clearly corresponds to a sum of two independent chi-square variates scaled by a factor of σ_ε^2. Thus, the methods described in this section have a firm basis in statistical theory.

5. Applications of the Rational Filter

The Wiener–Kolmogorov theory of signal extraction depends upon the availability of a statistical model which can represent the processes generating the data. The task of specifying the model can be approached in various ways.

Econometricians have often favoured a structural approach. This has two objectives. First, it is intended that the output of the model should mimic the data series as closely as possible. Secondly, it is proposed that the model should contain as many separate elements as there are discernible components in the data. Motivating this approach is the notion that the quality of the signal-extraction filter will be a function of the degree of realism in the underlying model. (A critical analysis of several examples of the structural approach has been provided by Garcia-Ferrer and Del Hoyo [9]).

An heuristic approach is one in which the model is determined solely with a view to ensuring that the resulting signal-extraction filter has certain preconceived properties. A common objective is to derive a lowpass filter with a designated cut-off frequency for which the transition from the pass band to the stop band is as rapid as possible, given the constraints of the filter order and the need to maintain numerical stability.

In this section, we shall concentrate on the heuristic approach which we shall illustrate in terms the Hodrick–Prescott [10] filter and the Butterworth filter, both of which have been used in the introduction to extract the trends from meteorological and economic data series. However, it should be emphasised that the methodology which has been outlined in the preceding sections is equally applicable to the structural approach to signal extraction which presupposes a realistic model.

The Hodrick–Prescott (H–P) filter, which is closely related to the Reinsch [14] smoothing spline, can be derived from the following model:

$$
\begin{aligned}
y(t) &= \xi(t) + \eta(t) \\
&= \frac{1}{(1-L)^2}\nu(t) + \eta(t).
\end{aligned} \tag{37}
$$

This equation represents a second-order random walk $\xi(t)$ which is obscured by disturbances or errors of observation which form a white-noise sequence $\eta(t)$. The equation can be written, alternatively, as

$$
(1-L)^2 y(t) = g(t) = \nu(t) + (1-L)^2\eta(t), \tag{38}
$$

which can be construed as an instance of equation (2). Notice, in particular, that the elements in this expression are stationary processes. The filter which is designed to extract $\nu(t)$ from $g(t)$ is

$$
\begin{aligned}
\beta_T(L) &= \frac{\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\eta^2(1-F)^2(1-L)^2} \\
&= \frac{1}{1+\lambda(1-F)^2(1-L)^2},
\end{aligned} \tag{39}
$$

where $\lambda = \sigma_\eta^2/\sigma_\nu^2$ is the so-called smoothing parameter. This filter will also serve to extract $\xi(t)$ from $y(t)$.

In comparing equation (39) with equation (4), it can be seen that $\delta_T(L) = 1$ and that $\delta_R(L) = (1-L)^2$. Also, $\gamma(z)\gamma(z^{-1}) = 1 + \lambda(1-z)^2(1-z^{-1})^2$ is the generating function which provides the coefficients of the Toeplitz matrix $GG' = \Omega_T + \lambda\Omega_R$ which is entailed in the finite-sample representation of the filter provided by equation (17). Here $\Omega_T = I$ is an identity matrix of order T.

The H–P filter has been used as a lowpass smoothing filter in numerous macroeconomic investigations (see for example, Hartley et. al. [11]) where it has been customary to set the smoothing parameter to certain conventional values. Thus, for example, the econometric computer package *Eviews 3.1* [8] imposes the following default values:

$$\lambda = \begin{cases} 100 & \text{for annual data,} \\ 1,600 & \text{for quarterly data,} \\ 14,400 & \text{for monthly data.} \end{cases} \tag{40}$$

In general, the H–P filter is characterised by a gradual transition from the pass band to the stop band. However, the default values have the effect of confining the pass band to a narrow margin of low frequencies; and this does ensure a rapid transition. With such values of λ, only the cycles with a duration of eight years or more are admitted to the tend. For lesser values of the smoothing parameter, the estimated trend is liable to comprise a broad range of components which acquire weights that decline slowly as the frequencies rise. Such heterogeneous combinations may be difficult to interpret.

In cases where equation (37) is a fair representation of the processes generating the data, it may be appropriate to take a structural approach to the H–P filter. Then the value of λ can be determined via the maximum-likelihood estimation of the model. In the case of the annual northern hemisphere temperature anomalies of Figure 1 (which have been analysed extensively by Bloomfield [3] and by Bloomfield and Nychka [4]), an estimated value of $\lambda = 650$ is obtained by fitting the model; and this has been used in generating the trend. Given that the residue sequence of Figure 4 resembles a white noise sequence, the approach seems justified in this case.

A filter which can be used to isolate a well-defined range of frequencies is the so-called Butterworth square-wave filter (See, for example, Roberts and Mullis [15]). The filter can be derived from an heuristic model represented by the equation

$$\begin{aligned} y(t) &= \xi(t) + \eta(t) \\ &= \frac{(1+L)^n}{(1-L)^d}\nu(t) + (1-L)^{n-d}\varepsilon(t). \end{aligned} \tag{41}$$

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

$$\beta_T(L) = \frac{(1+F)^n(1+L)^n}{(1+F)^n(1+L)^n + \lambda(1-F)^n(1-L)^n}, \tag{42}$$

where $\lambda = \sigma_\varepsilon^2/\sigma_\nu^2 = \{1/\tan(\omega_c)\}^{2n}$. Here ω_c represents a nominal cut-off frequency, which is the point where the gain functions of the lowpass filter

$\beta_T(L)$ and of the complementary highpass filter $\beta_R(L) = 1 - \beta_T(L)$ intersect at a value of 1/2. The finite-sample implementation of the filter is easily derived.

In common with that of the H–P filter, the rate of transition of the Butterworth filter increases as λ increases, which happens as the designated cut-off frequency ω_c in reduced. The rate of transition also increases with the filter order n. However, as n increases, the poles of the filter migrate toward the unit circle; and this can lead to problems of numerical stability.

The Butterworth filter can be used to separate two components of a data sequence which lie in disjoint frequency intervals. Ideally, there should be a dead space in the frequency spectrum of the data which separates the components. In that case, the parameters of the filter can be adjusted so that the transition from pass band to stop band takes place entirely within the dead space.

An example of the use of the Butterworth filter is provided by Figure 2 which depicts the data on the concentration of atmospheric carbon dioxide which come from the Mauna Loa observatory (see Boden et al. [5]). The data are monthly; and the objective in fitting a trend is to avoid incorporating any components of the annual cycle. At the same time, the trend should contain all cycles that have a period longer than one year. For example, the effects of major volcanic eruptions, which are liable to be transitory (lasting no longer that two years), ought be registered by the tend. The fitted trend in Figure 2 has been generated by selecting a filter order of $n = 6$ and cut-off frequency of $\omega_c = \pi/8$. The degree of differencing is $d = 2$. The residual sequence is depicted in Figure 5.

A further example of the use of the Butterworth filter is provided by Figure 3 which depicts the level of unemployment in Switzerland. The statistics, which are quarterly, run from the first quarter of 1980 through to the second quarter of of 1996.

The objective is to isolate the seasonal fluctuations which play on the back of the broad trend. This purpose is served by a Butterworth filter with an order of $n = 8$, a differencing of degree $d = 2$ and a nominal cut-off frequency of $\omega_c = 3\pi/8$. These choices place the transition band of the filter in a dead space which lies between the the seasonal components, which are concentrated at the frequency of $\pi/2$, and the trend components, which have frequencies that lie well below the cut-off point.

The seasonal fluctuations in Swiss unemployment are depicted in Figure 6. They have a remarkable regularity which is hardly apparent in the original data series depicted in Figure 3. Moreover, there is a modulation in the amplitude of these fluctuations which is correlated with the level of unemployment. This provides clear evidence of the phenomenon of labour hoarding whereby, in times of high economic activity, employers tend to retain the workers who might otherwise be dismissed in the lean season of the year.

It is notable that the H–P filter is incapable of separating the seasonal fluctuations in unemployment from the general trend. The reason is that it is not possible to adjust the filter so that its transition falls in the appropriate frequency band. The upshot is that the residual sequence, with which one might

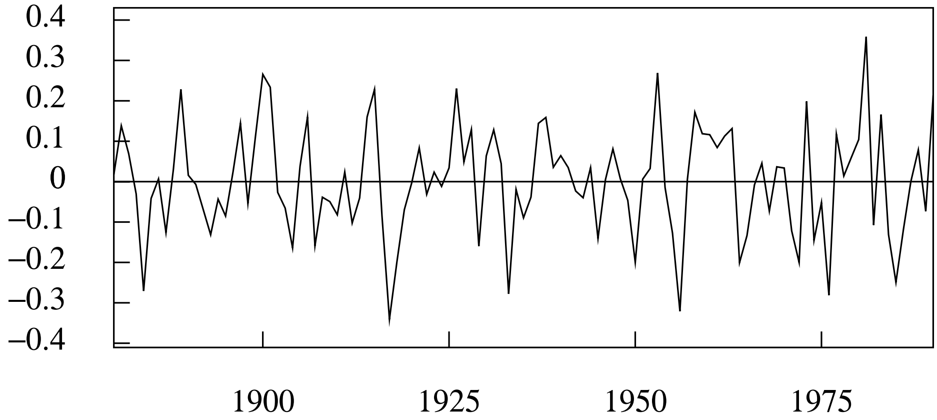

Figure 4.. The residual sequence from detrending the temperature data.

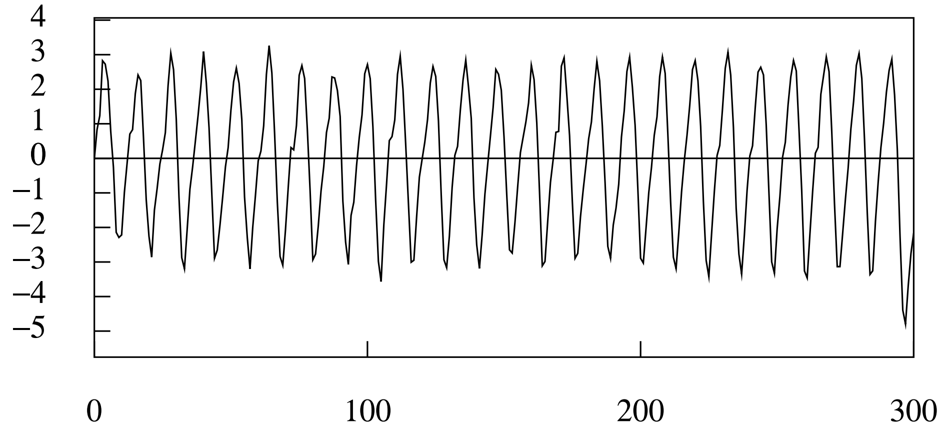

Figure 5. The annual fluctuations in the atmospheric concentration of carbon dioxide which surround the upward trend.

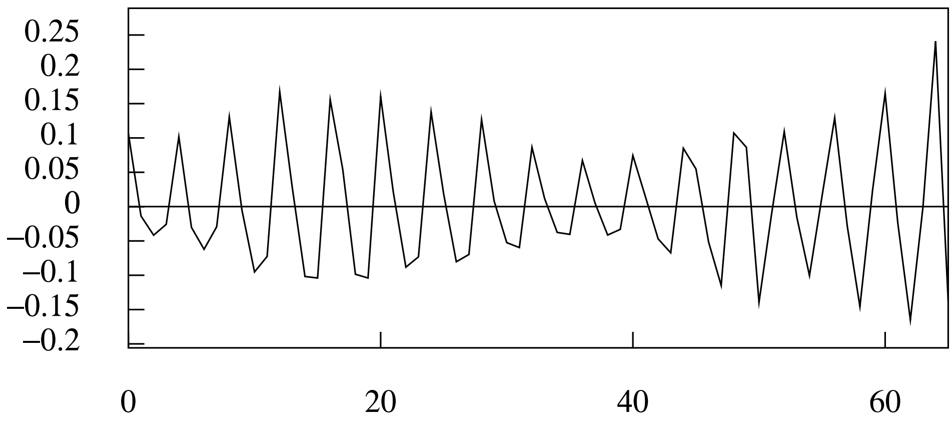

Figure 6. The residual sequence from detrending the Swiss unemployment figures, which gives evidence of labour hoarding.

propose to represent the seasonal effects, is bound to contain components that ought to be consigned to the trend.

Appendix: Orthogonal Projectors

Definition: Let P_1, P_2 be idempotent matrices of order $T \times T$ such that $P_1^2 = P_1$ and $P_2^2 = P_2$.

(a) We say that these matrices represent projectors which are mutually orthogonal in respect of the Σ^{-1} metric if $P_1 P_2 = 0$, $P_2 P_1 = 0$ and $P_1' \Sigma^{-1} P_2 = 0$.

(b) We say that the matrices represent complementary projectors if $P_1 + P_2 = I$.

Example: Let W and X be matrices of orders $T \times (T-k)$ and $T \times k$, respectively, which are complementary, such that $\text{Rank}[W, X] = T$, and mutually orthogonal in the Σ^{-1} metric, such that $W'\Sigma^{-1}X = 0$, where Σ is symmetric and positive definite. Define

$$P_W = W(W'\Sigma^{-1}W)^{-1}W'\Sigma^{-1} \quad \text{and} \quad P_X = X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}.$$

Then P_W and P_X are complementary projectors which are mutually orthogonal in respect of the Σ^{-1} metric.

The idempotency and the mutual orthogonality of these projectors is immediately discerned. We need only confirm their complementarity. Let $y = Wa + Xb$ be an arbitrary vector. Then

$$(P_W + P_X)y = P_W(Wa + Xb) + P_X(Wa + Xb) = Wa + Xb = y,$$

which shows that $P_W + P_X = I$.

Observe that P_X and $I - P_X$ are complementary projectors. Since the complement of P_X is uniquely defined, It follows that

$$P_W = I - P_X.$$

Now define $Q' = W'\Sigma^{-1}$. Then we can write

$$P_W = \Sigma Q(Q'\Sigma Q)^{-1}Q' = P_Q,$$

and we also have

$$P_Q = I - P_X.$$

The features of this example which are important for the text can be summarised as follows:

Lemma: Let Q, X be matrices of orders $T \times (T-k)$ and $T \times k$, respectively, which are complementary and mutually orthogonal in the ordinary sense such that $\text{Rank}[Q, X] = T$ and $Q'X = 0$. Let Σ be a positive definite matrix of order $T \times T$. Then there is the following matrix identity:

$$\Sigma Q(Q'\Sigma Q)^{-1}Q' = I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}.$$

Programs

The computer programs which have been used in implementing the methods described in this paper are available on a compact disc which is included in a book by the author: Pollock [13]

References

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

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

[3] Bloomfield, P., (1992), Trends in Global Temperature, *Climate Change,* **21,** 1–16.

[4] Bloomfield, P., and D. Nychka, (1992), Climate Spectra and Detecting Climate Change, *Climate Change,* **21,** 275–287.

[5] Boden, T.A., R.J. Sepanski and F.W. Stross., (1991), *Trends 91: A Compendium of Data on Global Change,* Carbon Dioxide Information Analysis Centre, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

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

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

[8] *Eviews 3.1.* QMS Quantitative Micro Software, 4521 Campus Drive Suite 336 Irvine, CA 92612-2699. (www.eviews.com).

[9] Garcia-Ferrer, A., and J. Del Hoyo, (1992), On Trend Extraction Models: Interpretation, Empirical Evidence and Forecasting Performance. *Journal of Forecasting*, **11** 645–665.

[10] Hodrick, R.J., and E.C. Prescott, (1997), Postwar U.S. Business Cycles: An Empirical Investigation, *Journal of Money, Credit, and Banking,* **29,** 1–16.

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

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

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

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

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

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

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