Statistical Signal Extraction and Filtering: A Partial Survey

D.S.G. POLLOCK

Department of Economics
University of Leiester

Abstract

A wide variety of techniques have been devised for the purpose of extracting the components of econometric data sequences. The traditional approach, which continues to prevail in the central statistical agencies of many countries, has relied upon filters that have been derived with refernce to the principles of actuarial graduation. Latterly, filters that are grounded in the Wiener–Kolmogorov theory of signal extraction have become prominent in academic research. There are several alternative ways of applying the Wiener–Kolmogorov filters to the data.

1 Introduction: The Semantics of Filtering

In common parlance, a filter is a device for removing solids or suspended particles from liquids. In the late 17th century, the term began to be used by the natural philosophers in a manner that gave expression to their understanding of the nature of light. It was recognised that white light is a compound of coloured lights of differing wavelengths. A coloured glass was seen as a device that selectively transmits some of the light, corresponding to a range of wavelengths, while blocking the remainder. Therefore, it was described as an optical filter.

A direct analogy with light led engineers, in the early 20th century, to talk of electronic filters. Electronic filters are constructed from capacitors, resistors and inductors. A circuit in which a voltage signal passes through an inductor, or in which a capacitor provides a path to earth, imposes less attenuation on low-frequency signals than on high-frequency signals. Therefore, it constitutes a lowpass filter. If the signal passes through a capacitor, or has a path to earth through an inductor, then the circuit imposes less attenuation on high-frequency signals than on low-frequency signals, and it constitutes highpass filter.

In these examples, one is imagining a stream or a current flowing continuously through the filter. The notion of a filter seems inappropriate to statistical time-series analysis, where the data are a sequence of discrete observations. However, over a period of half a century at least, there has been a gradual shift in electronic technology from analogue devices, which are naturally analysed in terms of continuous time, to digital devices, which are best described in terms of events occurring at discrete points in time. In the process, the terminology of electronic filtering has made a transition from the analogue to the digital domain; and electronic filtering has come to be known as signal processing.

Given the increasing commonality between digital signal processing and statistical time-series analysis, there are compelling reasons for why the two disciplines should share a common terminology, and this is what has transpired. Such is the convergence of these disciplines that, nowadays, their adherents contribute often to the same academic journals and they can be found at the same conferences.

Nevertheless, considerable differences remain, both of emphasis and of conceptualisation. In particular, statisticians tend to operate principally within the time domain, in which their discretely sampled data naturally reside, whereas engineers, who are familiar with harmonic motions and oscillating currents, feel at home in the frequency domain. The present account of linear filtering

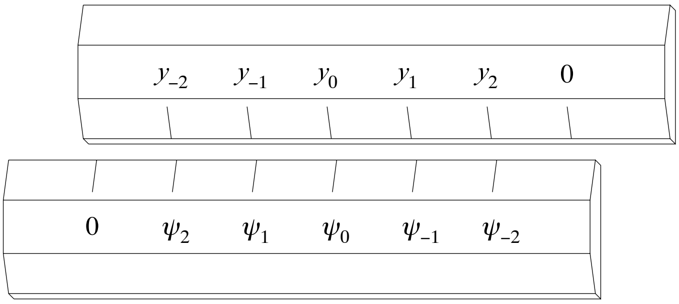

Figure 1: A method for finding the linear convolution of two sequences. The element $x_2 = \psi_0 y_2 + \psi_1 y_1 + \psi_2 y_0$ of the convolution may be formed by multiplying the adjacent elements on the two rulers and by summing their products.

and signal extraction has a statistical bias. It proceeds from the time domain to the frequency domain. It is also orientated towards econometric analysis, since econometrics is the primary discipline of the author.

Econometric data are supplied, in the main, by governmental agencies, such as the Central Statistical Office of the U.K. or the Bureau of the Census of the U.S. They come mainly at intervals of a year or a quarter, but there are also some monthly data. The fixity of these sampling rates has meant that, in the past, econometricians have not thought much about the effect of varying the rates at which the data are derived by sampling continuous processes. This is notwithstanding a venerable tradition of continuous-time econometrics, which adopts the premise that all processes should, in the first instance, be modelled in continuous time. The growth of financial econometrics and the advent of various theoretical developments, of which wavelet analysis is the principal one, have raised the issue of sampling rates anew, and we shall devote some attention to it.

2 Linear and Circular Convolutions

In the time-domain, a process of filtering corresponds to the convolution of two sequences. If $\psi(j) = \{\psi_j; j = 0 \pm 1, \pm 2, \ldots\}$ is the sequence of filter coefficients and if $y(t) = \{y_t; t = 0, \pm 1, \pm 2, \ldots\}$ is the data sequence, then the filtered sequence $\psi(j) * y(t) = x(t) = \{x_t; t = 0, \pm 1, \pm 2, \ldots\}$, which is the convolution of the two, has the generic element

$$x_t = \sum_j \psi_j y_{t-j} = \sum_j \psi_{t-j} y_j. \tag{1}$$

Adding the indices j and $t - j$ of the factors of the generic product of the RHS, gives the value t, which is the index of x_t on the LHS.

The process of convolution is also entailed in the multiplication of two polynomials or power series, since it is the process by which the coefficients of the product are obtained from those of its factors. By converting the sequences into series, one gains access to the algebra of polynomials and power series. We define the z-transforms of the sequences to be $\psi(z) = \sum_j \psi_j z^j$, $y(z) = \sum_t y_t z^t$ and $x(z) = \sum_t x_t z^t$. Thereafter, in place of (1), we may consider

$$x(z) = \psi(z) y(z). \tag{2}$$

Here, z is an algebraic indeterminate, which may be specified in a variety of useful ways. In particular, we may set $z = \exp\{-\mathrm{i}\omega\} = \cos(\omega) - \mathrm{i}\sin(\omega)$, where $\omega \in [0, 2\pi]$ is an angle measured in radians. This confines z to the circumference of the unit circle in the complex plane and, in the process, $\psi(\omega) = \psi(\exp\{-\mathrm{i}\omega\})$, $y(\omega) = y(\exp\{-\mathrm{i}\omega\})$ and $x(\omega) = x(\exp\{-\mathrm{i}\omega\})$ become objects within the frequency domain.

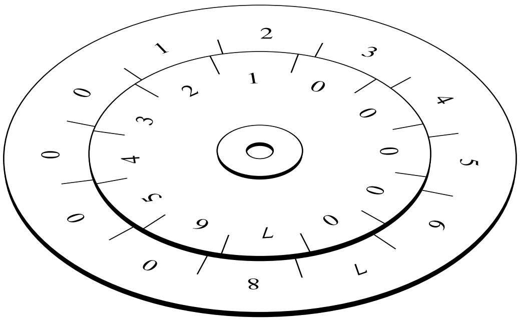

Figure 2: A device for finding the circular convolution of two sequences. The upper disc is rotated clockwise through successive angles of 30 degrees. Adjacent numbers on the two discs are multiplied and the products are summed to obtain the coefficients of the convolution.

Within the time domain, there are some alternative conceptualisations of the process of convolution that may prove helpful. The convolution of $\psi(j) = \{\psi_j; j = 0, \pm 1, \pm 2, \ldots\}$ and $y(t) = \{y_t; t = 0, \pm 1, \pm 2, \ldots\}$ entails the following products:

$$\begin{array}{ccccccc} \cdots & \psi_{-2}y_{-2} & \psi_{-1}y_{-2} & \psi_0 y_{-2} & \psi_1 y_{-2} & \psi_2 y_{-2} & \cdots \\ \cdots & \psi_{-2}y_{-1} & \psi_{-1}y_{-1} & \psi_0 y_{-1} & \psi_1 y_{-1} & \psi_2 y_{-1} & \cdots \\ \cdots & \psi_{-2}y_0 & \psi_{-1}y_0 & \psi_0 y_0 & \psi_1 y_0 & \psi_2 y_0 & \cdots \\ \cdots & \psi_{-2}y_1 & \psi_{-1}y_1 & \psi_0 y_1 & \psi_1 y_1 & \psi_2 y_1 & \cdots \\ \cdots & \psi_{-2}y_2 & \psi_{-1}y_2 & \psi_0 y_2 & \psi_1 y_2 & \psi_2 y_2 & \cdots \end{array} \tag{3}$$

The filtered sequence $x(t)$ is formed by summing the elements in each of the successive diagonals of the array that run in the SW–NE direction. Thus

$$\begin{array}{lc} x_{-4} & \cdots \psi_{-2}y_{-2} \cdots \\ x_{-3} & \cdots \psi_{-2}y_{-1} + \psi_{-1}y_{-2} \cdots \\ x_{-2} & \cdots \psi_{-2}y_0 + \psi_{-1}y_{-1} + \psi_0 y_{-2} \cdots \\ x_{-1} & \cdots \psi_{-2}y_1 + \psi_{-1}y_0 + \psi_0 y_{-1} + \psi_1 y_{-2} \cdots \\ x_0 & \cdots \psi_{-2}y_2 + \psi_{-1}y_1 + \psi_0 y_0 + \psi_1 y_{-1} + \psi_2 y_{-2} \cdots \\ x_1 & \cdots \psi_{-1}y_2 + \psi_0 y_1 + \psi_1 y_0 + \psi_2 y_{-1} \cdots \\ x_2 & \cdots \psi_0 y_2 + \psi_1 y_1 + \psi_2 y_0 \cdots \\ x_3 & \cdots \psi_1 y_2 + \psi_2 y_1 \cdots \\ x_4 & \cdots \psi_2 y_2 \cdots \end{array} \tag{4}$$

The first conceptualisation of this convolution entails what may be described as contragrade multiplication, which is also entailed by the concept of a moving average. It helps, in describing this, to consider two finite sequences. Imagine two rulers. One, denoted Y, bears the elements of the data sequence $\{y_{-2}, y_{-1}, y_0, y_1, y_2\}$. The other, denoted Ψ, bears the elements of the filter sequence in reverse: $\{\psi_2, \psi_1, \psi_0, \psi_{-1}, \psi_{-2}\}$. These are shown in Figure 1. The two rulers approach each other from opposite directions: Ψ from the left and Y from the right.

When the rulers first meet, the product $x_{-4} = \psi_{-2}y_{-2}$ is formed and recorded. Then, the rulers take a contragrade step which brings ψ_{-2} adjacent to y_{-1} and ψ_{-1} adjacent to y_{-2}. The products of these adjacent elements are formed and added to give $x_{-3} = \psi_{-2}y_{-1} + \psi_{-1}y_{-2}$. A further contragrade step is taken and the product $x_{-2} = \psi_{-2}y_0 + \psi_{-1}y_{-1} + \psi_0 y_{-2}$ is formed. Successive steps are taken and the products are formed until none of the nonzero elements of Y and Ψ are adjacent.

This is linear convolution. There is no necessity for the sequences $\psi(j)$ and $y(t)$ to be finite. However, if they are infinite sequences, then a sufficient condition for the elements of their convolution product to be finite-valued is that both sequences should have elements that are bounded in value and that the elements of the filter sequence should be absolutely summable.

There is also a process of circular convolution, which is applicable to finite sequences. If these are $\{\psi_0, \psi_1, \ldots, \psi_n\}$ and $\{y_0, y_1, \ldots, y_n\}$, then the generic element of their circular convolution is

$$x_t^\circ = \sum_j \psi_j^\circ y_{t-j}^\circ = \sum_j \psi_{t-j}^\circ y_j^\circ, \tag{5}$$

wherein $\psi_j^\circ = \psi_{j \bmod n}$ and $y_t^\circ = y_{t \bmod n}$.

For an analogy of the process of circular convolution, one can imagine two discs placed one above the other on a common axis, with the rim of the lower disc protruding. The device is shown in Figure 1. On this rim, are written the elements of the sequence $\{y_0, y_1, \ldots, y_{n-1}\}$ at equally spaced intervals in a clockwise order. On the rim of the upper disc, are written the elements of $\{\psi_0, \psi_1, \ldots, \psi_n\}$ equally spaced in an anticlockwise order. At the start of the process of circular convolution, ψ_0 and y_0 are in alignment, and the pairs $(\psi_0, y_0), (\psi_1, y_{n-1})$, $\ldots, (\psi_{n-1}, y_1)$ are read from the disc and added to give x_0°. Then, the upper disc is turned clockwise through an angle of $2\pi/n$ radians and the pairs $(\psi_0, y_1), (\psi_1, y_0), \ldots, (\psi_{n-1}, y_2)$ are read from the disc and added to give x_1°. The process continues until the $(n-1)$th turn when the pairs $(\psi_0, y_{n-1}), (\psi_1, y_{n-2}), \ldots, (\psi_{n-1}, y_0)$ give rise to x_{n-1}°. One more turn of the disc would bring us back to the starting position, wherafter we could begin to generate a repetition of the sequence $\{x_0^\circ, x_1^\circ, \ldots, x_{n-1}^\circ\}$.

2.1 Kernel Smoothing

The second conceptualisation of the convolution operation may be described as kernel multiplication. Let $y(t) = \{y_t; t = 0, \pm 1, \pm 2, \ldots\}$ be a sequence of indefinite length, and let $\psi(j) = \{\psi_j; j = 0, \pm 1, \pm 2, \ldots\}$ be a sequence of finite length, or at least one of which the absolute values of the elements sum to a finite value. The latter sequence is a so-called kernel function or filter, denoted Ψ.

When Ψ encounters y_0, all of its elements are multiplied by that value. Thereafter, they are accumulated in the registers of the derived sequence $x(t) = \{x_t; t = 0, \pm 1, \pm 2, \ldots\}$. Thus, on considering the middle row of (3), we see that $y_0\psi_{-2}$ is accumulated to x_{-2}, $y_0\psi_{-1}$ is accumulated to x_{-1}, $y_0\psi_0$ is accumulated to x_0, and so on. When this process is ended, Ψ is moved to the right were it encounters y_1. Then, $y_1\psi_{-2}$ is accumulated to x_{-1}, $y_1\psi_{-1}$ is accumulated to x_0, $y_1\psi_0$ is accumulated to x_1, and so on.

If the elements of Ψ sum to unity, and if its profile resembles that of a probability mass function, then the process that we have described can be regarded as a smoothing operation, whereby each element of Y is dispersed over a range of neighbouring points as the filter or kernel Ψ passes along the sequence. The condition $\sum_j \psi_j = 1$ that the kernel elements sum to unity can be expressed in terms of the z-transform $\psi(z)$ as $\psi(1) = 0$. Observe that the condition implies that the weights associated with the sample values that are accumulated to x_k will also sum to unity, for the reason that the weights are the kernel elements.

The concept of kernel smoothing is central to the theories of density function estimation and nonparametric regression. In these contexts, the kernel Ψ typically becomes a continuous function, which is effective in distributing the mass of a discrete observation over an interval of the real line.

The kernel function is often a probability density function or mass function, which is symmetric with a zero mean and a finite variance. In that case, the standard deviation become a scaling factor, which governs the dispersion of the kernel and hence the extent to which it smoothes the data.

However, it is unnecessary to restrict the class of kernel functions in this way. The restriction that the kernel weights should sum to unity constrains the corresponding filter to be a lowpass filter that preserves all elements in the vicinity of zero frequency. If $\psi(z)$ is the z-transform of a lowpass filter, then $1 - \psi(z)$ is the z-transform of the complementary highpass filter. A highpass filter, which is intended to remove the low-frequency trend from the data, should be subject to the restriction that its coefficients should sum to zero.

According to a common terminology, which is somewhat ambiguous, the scaling factor of the kernel is described as its bandwidth. In one perception, the band in question is the neighbourhood

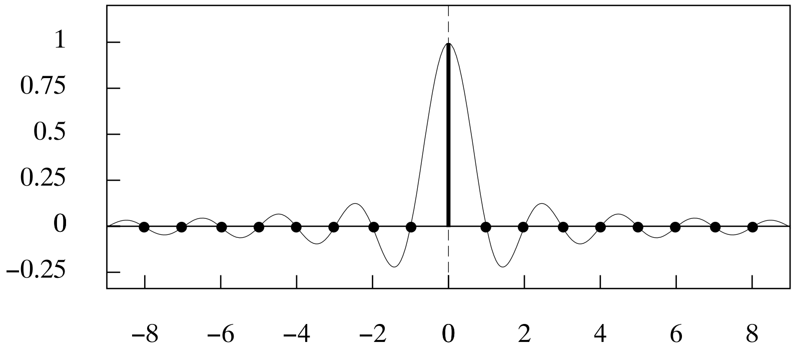

Figure 3: The sinc function $\psi(t) = \sin(\pi t)/\pi t$.

of a data point. In that case, the scaling factor governs the width of the support of the kernel function, on the understanding that it is finite.

According to an alternative interpretation, the bandwidth refers to the range of frequencies in the spectral decomposition of the kernel. This usage accords with the popular understanding of frequency-related phenomena, which has been fostered by the widespread availability of domestic electronic appliances. These alternative interpretations are closely linked. In particular, a narrow bandwidth in the time domain implies a wide bandwidth in the frequency domain and vice versa.

Continuous kernel functions can be used to reconstitute a continuous function of time from regularly sampled observations. The Shannon–Nyquist theorem, which we shall propound later, indicates that, if the sinusoidal elements of which a stationary time series is composed are band-limited to the frequency interval $[0, \pi]$, which is to say that there is no element that completes its cycle in less than two sample periods, then a perfect reconstruction of the underlying function can be obtained from its sampled values using the sinc function $\psi(t) = \sin(\pi t)/\pi t$ of Figure 3 as the kernel smoother. In this case, it is entirely accurate to say that the sinc function has a frequency bandwidth of π radians.

The sinc function has a value of unity at $t = 0$ and a value of zero on all other integer points. Thus, instead of distributing the values of the data points over other adjacent integers, the sinc function leaves those values intact; and it adds nothing to the other integers. However, it does attribute values to the non-integer points that lie in the interstices, thereby producing a continuous function from discrete data. Moreover, the data can be recovered precisely by sampling the continuous function at the integer points.

3 Local Polynomial Regression

One way of estimating the trend in a sequence $\{y_t; t = 0, 1, \ldots, T-1\}$ is to interpolate through the data a polynomial in the powers of the time index t. However, there can be disadvantages in representing the trend via an analytic function. Such a function is completely determined by the values of its derivatives at any point in its domain; and any local features of the data that are captured by the function will also have global effects.

The characteristics of the trend in the locality of y_t will be reflected more effectively in a polynomial fitted to a limited set of adjacent data values $\{y_{t-j}; j = 0, \pm 1, \ldots, \pm m\}$. The resulting local polynomial, which may be denoted by $\gamma(j) = \gamma_0 + \gamma_1 j + \cdots + \gamma_p j^p$, will comprise powers of the index j. Its central value at $j = 0$, which is $x_t = \gamma(0) = \gamma_0$, will provide an estimate of the trend at time t. A sequence of such local polynomials, fitted to the points within a window that moves through the data in step with t, will provide a sequence of trend estimates.

The polynomials may be fitted by minimising a weighted sum of squares of the deviations from the local data:

$$S(t)=\sum_{j=-m}^{m}\frac{1}{\lambda_j}\big\{y_{t-j}-\gamma(j)\big\}^2. \tag{6}$$

Then, the estimates of the polynomial coefficients will be linear functions of these data values. In particular, the minimisation of $S(t)$ will determine a set of moving-average coefficients $\{\psi_j; j = 0, \pm1, \ldots, \pm m\}$ such that $x_t = \sum \psi_j y_{t-j}$. These coefficients are invariant with respect to the location of the data window; and, therefore, they serve to provide smoothed values throughout the sample, with the exception of the first m sample points and the last m sample points, which demand some special treatment.

To examine this method in more detail, let us define

$$P=\begin{bmatrix} 1 & -m & m^2 & \ldots & (-m)^p \\ 1 & 1-m & (1-m)^2 & \ldots & (1-m)^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & m-1 & (m-1)^2 & \ldots & (m-1)^p \\ 1 & m & m^2 & \ldots & m^p \end{bmatrix} = \begin{bmatrix} p'_{-m} \\ p'_{1-m} \\ \vdots \\ p'_0 \\ \vdots \\ p'_{m-1} \\ p'_m \end{bmatrix}, \tag{7}$$

which is the matrix of a basis for the local polynomial, together with the diagonal weighting matrix

$$\Lambda = \mathrm{diag}\{\lambda_{-m}, \lambda_{1-m}, \ldots, \lambda_{m-1}, \lambda_m\}. \tag{8}$$

Then, the vector $\gamma = [\gamma_0, \gamma_1, \ldots, \gamma_p]'$ of the coefficients of the polynomial $\gamma(j)$ that minimises $S(t)$ is obtained as the solution to the following normal equations of the local polynomial regression:

$$P'\Lambda^{-1}P\gamma = P'\Lambda^{-1}y. \tag{9}$$

Here,

$$y = [y_{t-m}, y_{t+1-m}, \ldots, y_t, \ldots, y_{t+m-1}, y_{t+m}]' \tag{10}$$

is the vector of the observations in the vicinity of y_t that lie within the data window. The smoothed value to replace y_t is

$$\begin{aligned}\gamma_0 = p'_0\gamma &= p'_0(P'\Lambda^{-1}P)^{-1}P'\Lambda^{-1}y \\ &= \psi' y,\end{aligned} \tag{11}$$

where $p'_0 = [1, 0, \ldots, 0]$ is the central row of the matrix P of (7); and it is manifest that the filter weights in $\psi = [\psi_{-m}, \ldots, \psi_m]'$ do not vary as the filter passes through the sample. Also, it can be seen that $\psi'P = p'_0(P'\Lambda^{-1}P)^{-1}P'\Lambda^{-1}P = p'_0$, which is to say that

$$\sum_{j=-m}^{m}\psi_j = 1, \quad \text{and} \quad \sum_{j=-m}^{m}\psi_j j^n = 0 \quad \text{for} \quad n = 1, \ldots, p. \tag{12}$$

These conditions are necessary and sufficient to ensure that $\sum_{j=-m}^{m}\psi_j\gamma(j) = \gamma_0$. The consequence is that the filter will transmit, without alteration, not only the ordinates of $\gamma(j)$ sampled at the integer points but also those of any other polynomial of degree p or less.

Also observe that, by projecting the local data vector on the polynomial basis provided by P, we would obtain a vector

$$\hat{y} = P(P'\Lambda^{-1}P)^{-1}P'\Lambda^{-1}y \tag{13}$$

comprising a set of $2m + 1$ smoothed values to replace those of y. In fact, we chose to select from this vector only the central value, denoted by x_t, which becomes the replacement for y_t. With this value in hand, the data window can be moved forwards, which is a matter of deleting the element y_{t-m} from one end of the vector y and appending a new data value y_{t+m+1} to the other end. Then, another smoothed value can be generated to replace y_{t+1}.

We must also consider the circumstances that arise when the data window reaches the end of the sample $y_0, \ldots, y_{T-1}$ and can move no further, which is when $t + m = T - 1$. Then, $x_t = \hat{y}_t$ is available as the central value of $\hat{y}$, whereas the smoothed values $x_{t+1}, \ldots, x_{T-1}$, which would otherwise depend upon extra-sample data values, are available from $\hat{y}$ as the succeeding elements $\hat{y}_{t+1}, \ldots, \hat{y}_{T-1}$. Under this construction, the final $m + 1$ smoothed values are generated according to the formulae

$$\begin{aligned} x_{t+i} = \hat{y}_{t+i} = \psi_i' y = p_i'(P'\Lambda^{-1}P)^{-1}P'\Lambda^{-1}y, \\ \text{where} \quad t = T - m - 1 \quad \text{and} \quad i = 0, \ldots, m, \end{aligned} \tag{14}$$

and where p_i' is the ith row below the central row p_0' of the matrix P.

Thus, there is a filter, with coefficients in the vector $\psi = \psi_0$, that is applicable to points in the middle of the sample, and there is a set of auxiliary filters, with coefficients in vectors $\psi_i; i = 1, \ldots, m$, that are applicable at the ends of the sample. The auxiliary filters at the upper end of the sample all comprise the same set of $2m + 1$ points, which lie within the data window when its progress through the sample is halted.

In an alternative method, the data window continues to move through the sample after the point $t = T - m - 1$ has been reached. From then onwards, its length contracts, through a diminishing number of points ahead of t, until $t = T - 1$. Then, the window, which comprises the $m + 1$ points $y_{T-m-2}, \ldots, y_{T-1}$, no longer looks forwards in time.

Let $t > T - m - 1$ be the current index. Then, the points falling within the data window are contained in the vector $y_1 = [y_{t-m}, \ldots, y_t \ldots, y_{T-1}]'$. The corresponding rows of the matrix P are in $P_1 = [p_{-m}, \ldots, p_0, \ldots, p_k]'$. The remaining rows in $P_2 = [p_{k+1}, \ldots, p_m]'$ are to be disregarded. The smoothed value to replace y_t is provided by

$$x_t = p_0'(P_1'\Lambda_1^{-1}P_1)^{-1}P_1'\Lambda_1^{-1}y_1 = \psi_q' y_1, \tag{15}$$

where q is the number of points that have been lost from the upper half of the data window. Then, $\psi_0 = \psi$ continues to denote the vector of the coefficients of a filter that can be applied to points in the middle of the sample, whereas $\psi_q; q = 1, \ldots, m$ denote the coefficient vectors of a sequence of filters of diminishing length that can be applied to the points at the end of the sample.

An alternative way of enabling the filter to reach the end of the sample is to represent the requisite extra-sample values by their predictions. Thereafter, one can apply the symmetric moving average to a data sequence comprising both sample values and predictions. One recourse is to generate the predictions via extrapolations of the local polynomial fitted to the sample values that lie within the current data window when $t > T - m - 1$. These predictions are provided by the vector

$$\hat{y}_2 = P_2(P_1'\Lambda_1^{-1}P_1)^{-1}P_1'\Lambda_1^{-1}y_1. \tag{16}$$

Define $M_1 = P_1'\Lambda_1^{-1}P_1$ and $M_2 = P_2'\Lambda_2^{-1}P_2$ such that $M_1 + M_2 = P'\Lambda^{-1}P$. Also let $\hat{y}' = [y_1', \hat{y}_2']$. Then, the smoothed value that incorporates the predictions is provided by

$$\begin{aligned} x_t &= p_0'(P'\Lambda^{-1}P)^{-1}P'\Lambda^{-1}\hat{y} \\ &= p_0'(M_1 + M_2)^{-1}(I + M_2M_1^{-1})P_1'\Lambda_1^{-1}y_1. \end{aligned} \tag{17}$$

Now observe that

$$I + M_2M_1^{-1} = (M_1 + M_2)M_1^{-1} \quad \text{implies} \quad (M_1 + M_2)^{-1}(I + M_2M_1^{-1}) = M_1^{-1}.$$

Therefore, (17) delivers $x_t = p_0'M_1^{-1}P_1'\Lambda_1^{-1}y_1$, which is none other than the value that is indicated by (15).

This result is due to Wallis (1981). It suggests that, in principle, a procedure based on a time-invariant filter that overcomes the end-of-sample problem by using predictions based on sample values can be replaced by an equivalent procedure that applies a time-varying filter to the sample points alone.

The technique of filtering via local polynomial regression becomes fully specified only when the regression weights within Λ are determined. The matter is dealt with in the following example. An account of local polynomial regression in a wider context than the present one has been provided by Proietti and Luati (2006). Other sources are Fan and Gijbels (2002) and Simonoff (1996).

Example. The requirement of Henderson (1916) was for a symmetric filter that would transmit a cubic polynomial time trend without distortion. It was also required that the filtered sequence should be as smooth as possible.

Consider the normal equations of (9) in the case where the polynomial degree is $p = 3$. The generic element in the rth row and kth column of the matrix $P'\Lambda^{-1}P$ is $\sum_{j=-m}^{m} w_j j^{r+k}$ $= s_{r+k}$ where, for notational convenience, we are using $w_j = \lambda_j^{-1}$. The filter will be symmetric if and only if the regression weights are symmetric such that $w_j = w_{-j}$ and, under these conditions, it follows that $s_{r+k} = 0$ if $r + k$ is odd. Therefore, the normal equations take the form of

$$\begin{bmatrix} s_0 & 0 & s_2 & 0 \\ 0 & s_2 & 0 & s_4 \\ s_2 & 0 & s_4 & 0 \\ 0 & s_4 & 0 & s_6 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} = \begin{bmatrix} \sum w_j y_{t-j} \\ \sum j w_j y_{t-j} \\ \sum j^2 w_j y_{t-j} \\ \sum j^3 w_j y_{t-j} \end{bmatrix}. \tag{18}$$

Only the first and the third of these equations are involved in the determination of γ_0 via

$$\begin{bmatrix} \gamma_0 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} s_0 & s_2 \\ s_2 & s_4 \end{bmatrix}^{-1} \begin{bmatrix} \sum w_j y_{t-j} \\ \sum j^2 w_j y_{t-j} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \sum w_j y_{t-j} \\ \sum j^2 w_j y_{t-j} \end{bmatrix}. \tag{19}$$

Thus

$$\gamma_0 = \sum_{j=-m}^{m} \psi_j y_{t-j} = \sum_{j=-m}^{m} (a + bj^2) w_j y_{t-j}, \tag{20}$$

where

$$as_0 + bs_2 = 1 \qquad \text{and} \qquad as_2 + bs_4 = 0. \tag{21}$$

It is now a matter of determining the filter coefficients $\psi_j = (a + bj^2)w_j$ in accordance with the smoothness criterion.

The criterion adopted by Henderson was that the variance of the third differences of the filtered sequence should be at a minimum. This requires that the process generating the sequence should be specified sufficiently for the variance to be defined. An appropriate model is one in which the data are generated by a cubic polynomial with added white noise: $y(t) = \beta(t) + \varepsilon(t)$.

The (forward) difference operator Δ has the effect that $\Delta y(t) = y(t + 1) - y(t)$. Also, the third difference of a cubic polynomial is some constant c. Therefore, $\Delta^3 y(t) = c + \Delta^3\varepsilon(t)$; and, if $V(\varepsilon_t) = \sigma^2$, it follow that

$$V\{\Delta^3 y(t)\} = \sum_j \{\Delta^3\psi(j)\}^2\sigma^2, \tag{22}$$

where $\psi(j) = \{\psi; j = 0, \pm1, \pm2, \ldots\}$ denotes the indefinite extension of the sequence of filter coefficients, formed by supplementing them with the set of zero-valued elements $\{\psi_{m+j} = 0; j = \pm1, \pm2, \ldots\}$. (The notation $\Delta^3\psi(j)$ recognises the fact that Δ operates upon infinite sequences. The usual notation of econometrics suggests that it is applicable to isolated elements of a sequence, but this is misleading.)

The resulting criterion is

$$\text{Minimise} \quad \sum_j \{\Delta^3\psi(j)\}^2 \quad \text{subject to} \quad \sum_j \psi_j = 1, \quad \sum_j j^2\psi_j = 0. \tag{23}$$

The two side conditions are from (12). The remaining conditions of (12), which are $\sum_j j\psi_j = \sum_j j^3\psi_j = 0$ are satisfied automatically in consequence of the symmetry about ψ_0 of the sequence $\psi(j)$. The constrained minimisation, which can be achieved using Lagragean multipliers, indicates that

$$\Delta^6\psi(j - 3) = a + bj^2, \quad \text{for} \quad j = 0, \pm1, \ldots, \pm m. \tag{24}$$

This implies that the filter coefficients are the ordinates of a polynomial in j of degree 8, namely $\psi(j) = \delta(j)(a + bj^2)$, of which the 6th difference is the quadratic function $a + bj^2$. For condition of (24) to be satisfied, it is necessary that $\psi(j)$ should be specified for the additional values of $j = \pm(m+1), \pm(m+2), \pm(m+3)$. Here, the ordinates are all zeros. It follows that the polynomial, which must have zeros at these six points, must take the form of

$$\psi(j) = \{(m+1)^2 - j^2\}\{(m+2)^2 - j^2\}\{(m+3)^2 - j^2\}(a + bj^2). \tag{25}$$

There are only two remaining parameters to be determined, which are a and b. They are determined via the conditions of (21).

Kenny and Durbin (1982) have found that

$$\psi_j \propto \{(m+1)^2 - j^2\}\{(m+2)^2 - j^2\}\{(m+3)^2 - j^2\}(3(m+2)^2 - 16 - 11j^2), \tag{26}$$

where the constant of proportionality is chosen to ensure that the coefficients sum to unity.

The Henderson filters require to be supplemented by asymmetric filters designed to cope with the end-of-sample problem. It might seem appropriate to use the techniques that have been described in the text prior to this section. They would entail the extrapolation of a local cubic model of the trend. However, experience has shown that it is more appropriate to depend upon a linear extrapolation.

A linear extrapolation is the basis of the so-called Musgrave (1964a, b) filters that have long been used in central statistical agencies in conjunction with the Henderson filters. Doherty (2001) has given an account of the origin of these filters and of the theory that lies behind them. Gray and Thomson (2002) have provided an exhaustive treatment of the theory of end-of-sample filters.

The Henderson filters have played a dominant *role* in the methods of trend estimation and seasonal adjustment that have been deployed in common by numerous central statistical agencies. Recently, there have been indications that they may be ceding their place to other filtering methods, such as those that are described in section 10 of this paper, which use ARIMA models to describe the components of the data. (See, for example, Monsell *et al.* (2003).) In particular, the TRAMO–SEATS program of Caporello and Maravall (2004) has attracted the widespread attention amongst the statistical agencies. It has been implemented in conjunction with the X-12-ARIMA program of the U.S. Bureau of the Census in the *Demetra* program of Statistical Office of the European Commission—see Eurostat (2002).

4 The Concepts of the Frequency Domain

According to the basic result of Fourier analysis, it is always possible to approximate an arbitrary function, defined over a finite interval of the real line and having a finite number of discontinuities therein, by a weighted sum of sine and cosine functions of harmonically increasing frequencies.

Similar results apply in the case of sequences, which may be regarded as functions mapping from the set of integers onto the real line. For a sample of $T = 2n$ observations $y_0, y_1, \ldots, y_{T-1}$, it is possible to devise an expression of the form

$$\begin{aligned} y_t &= \sum_{j=0}^{n} \rho_j \cos(\omega_j t - \theta_j) \qquad (27) \\ &= \sum_{j=0}^{n} \Big\{ \alpha_j \cos(\omega_j t) + \beta_j \sin(\omega_j t) \Big\}, \end{aligned}$$

wherein $\omega_j = 2\pi j/T$ is a multiple of the fundamental frequency $\omega_1 = 2\pi/T$. Here, in the second expression, there are $\alpha_j = \rho_j \cos(\theta_j)$ and $\beta_j = \rho_j \sin(\theta_j)$. Squaring and adding these gives $\rho_j^2 = \alpha_j^2 + \beta_j^2$. The equality of (27) follows in view of the trigonometrical identity

$$\cos(A - B) = \cos(A)\cos(B) + \sin(A)\sin(B). \tag{28}$$

Thus, the elements of a finite sequence can be expressed exactly in terms of a finite number of sines and cosines. A continuous function that interpolates the elements of the sequence can be obtained by replacing the integer-valued time index t by an argument that varies continuously.

The sequence $y(t) = \{y_t; t = 0, \pm 1, \pm 2, \ldots\}$, expressed in the manner of (27), is periodic with a period T equal to the length of the sample. If we confine our attention to a segment of length T, then the periodicity will not be evident. However, we shall also have occasion to consider the periodic extension of the sample, obtained be replicating sample elements over all preceding and succeeding intervals of T points, which is denoted by $y(t)$.

We may observe that, within (27), there are $\sin(\omega_0 t) = \sin(0) = 0$ and $\sin(\omega_n t) = \sin(\pi t) = 0$. Therefore, disregarding these zero-valued functions, there are as many trigonometrical basis functions in the sum as there are observations in the data sequence $\{y_0, y_1, \ldots, y_{T-1}\}$. Thus, the so-called Fourier coefficients

$$\{\alpha_0, \alpha_1, \beta_1, \ldots, \alpha_{n-1}, \beta_{n-1}, \alpha_n\},$$

which are obtained by projecting the data sequence onto the trigonometrical basis, provide a complete summary of the sampled information.

Since the trigonometrical functions are mutually orthogonal, the Fourier coefficients can be obtained via a set of T simple inner-product formulae, which are in the form of ordinary univariate least-squares regressions, with the values of the sine and cosine functions at the points $t = 0, 1, \ldots, T-1$ as the regressors. Let $c_j = [c_{0,j}, \ldots, c_{T-1,j}]'$ and $s_j = [s_{0,j}, \ldots, s_{T-1,j}]'$ represent vectors of T values of the generic functions $\cos(\omega_j t)$ and $\sin(\omega_j t)$ respectively, and let $y = [y_0, \ldots, y_{T-1}]'$ be the vector of the sample data and $\iota = [1, \ldots, 1]'$ a vector of units. Then, the 'regression' formulae for the Fourier coefficients are

$$\alpha_0 = (\iota'\iota)^{-1}\iota' y = \frac{1}{T}\sum_t y_t = \bar{y}, \tag{29}$$

$$\alpha_j = (c_j'c_j)^{-1}c_j' y = \frac{2}{T}\sum_t y_t \cos\omega_i t, \tag{30}$$

$$\beta_j = (s_j's_j)^{-1}s_j' y = \frac{2}{T}\sum_t y_t \sin\omega_j t, \tag{31}$$

$$\alpha_n = (c_n'c_n)^{-1}c_j' y = \frac{1}{T}\sum_t (-1)^t y_t. \tag{32}$$

However, in calculating the coefficients, it is more efficient to use the family of specialised algorithms known as fast Fourier transforms, which deliver the spectral ordinates from which the Fourier coefficients are obtained directly.

Equation (27) can be written in a more concise manner using the Euler equations:

$$\cos(\omega_j t) = \frac{1}{2}(e^{\mathrm{i}\omega_j t} + e^{-\mathrm{i}\omega_j t}) \quad \text{and} \quad \sin(\omega_j t) = \frac{-\mathrm{i}}{2}(e^{\mathrm{i}\omega_j t} - e^{-\mathrm{i}\omega_j t}). \tag{33}$$

Then,

$$\begin{aligned} y_t &= \sum_{j=0}^{n}\Big(\frac{\alpha_j - \mathrm{i}\beta_j}{2}\Big)e^{\mathrm{i}\omega_j t} + \sum_{j=0}^{n}\Big(\frac{\alpha_j + \mathrm{i}\beta_j}{2}\Big)e^{-\mathrm{i}\omega_j t} \\ &= \sum_{j=0}^{n}\zeta_j e^{\mathrm{i}\omega_j t} + \sum_{j=0}^{n}\zeta_j^* e^{-\mathrm{i}\omega_j t} = \sum_{j=-n}^{n}\zeta_j e^{\mathrm{i}\omega_j t}, \end{aligned} \tag{34}$$

where $\zeta_j = (\alpha_j - \mathrm{i}\beta_j)/2$, which has $\zeta_j^* = \zeta_{-j} = (\alpha_j + \mathrm{i}\beta_j)/2$ as its complex conjugate. Also $\zeta_0 = \alpha_0$ and $\zeta_n = \alpha_n$.

The exponential $\exp(\mathrm{i}\omega_j) = \exp(\mathrm{i}2\pi j/T)$ is T-periodic in the index j. Therefore, $\exp(\mathrm{i}\omega_{-j}) = \exp(\mathrm{i}\omega_{T-j})$ and, by taking $\zeta_j^* = \zeta_{-j} = \zeta_{T-j}$, we may write

$$y_t = \sum_{j=0}^{T-1}\zeta_j e^{\mathrm{i}\omega_j t}, \tag{35}$$

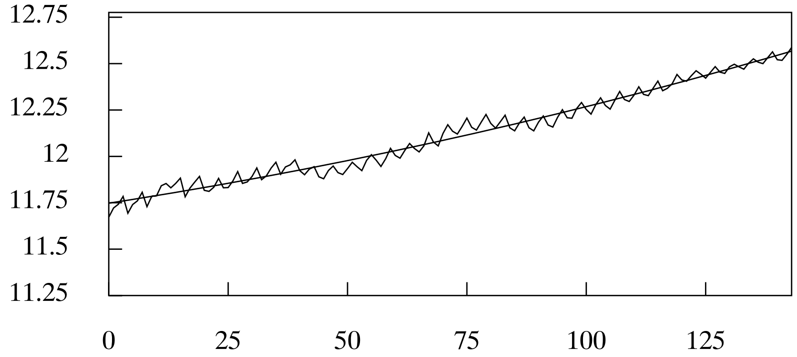

Figure 4: The quarterly sequence of the logarithms of the GDP in the U.K. for the years 1970 to 2005, inclusive, together with a quadratic trend interpolated by least squares regression.

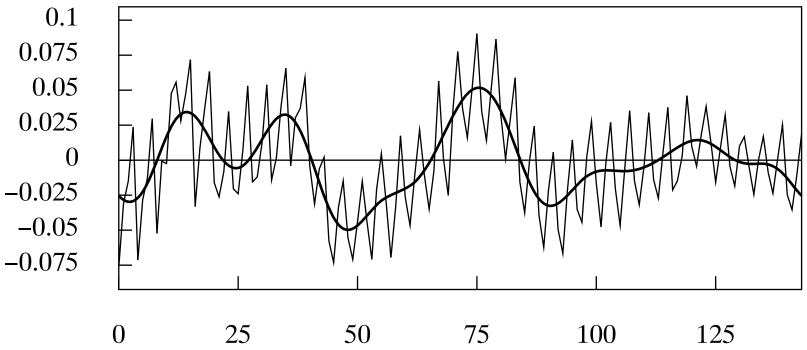

Figure 5: The residual sequence from fitting a quadratic trend to the income data of Figure 4. The interpolated line represents the business cycle

wherein the frequency index $j = 0, 1, \ldots, T-1$ has the same range as the temporal index t. The sequence $\zeta_0, \zeta_1, \ldots, \zeta_{T-1}$ constitutes the spectral ordinates of the data. The inverse of (35) is the transform that maps from the data to the spectral ordinates:

$$\zeta_j = \frac{1}{T}\sum_{t=0}^{T-1} y_t e^{-\mathrm{i}\omega_j t}. \tag{36}$$

The expression $\zeta_j = (\alpha_j - \mathrm{i}\beta_j)/2$ is recovered by using the identity $\exp\{-\mathrm{i}\omega_j t\} = \cos(\omega_j t) - \mathrm{i}\sin(\omega_j t)$ together with the equations (30) and (31) for α_j and β_j. Equations (35) and (36) together summarise the discrete Fourier transform.

4.1 The Periodogram

The power of a sequence is synonymous with the mean-square deviation which, in statistical terms, is its variance. The power of the sequence $x(t) = \rho_j \cos(\omega_j)$ is $\rho_j^2/2$. This result can be obtained in view of the identity $\cos^2(\omega_j t) = \{1 + \cos(2\omega_j t)\}/2$, for the average of $\cos(2\omega_j t)$ over an integral number of cycle is zero. The assemblage of values $\rho_1^2/2, \ldots, \rho_n^2/2$ constitutes the power spectrum of $y(t)$, which becomes the periodogram when scaled by a factor T. Their sum equals the variance of the sequence:

$$\frac{1}{T}\sum_{t=0}^{T-1}(y_t - \bar{y})^2 = \frac{1}{2}\sum_{j=1}^{n-1}\rho_j^2 + \alpha_n^2. \tag{37}$$

The periodogram is effective in revealing the spectral structure of the data and in guiding the business of extracting its components.

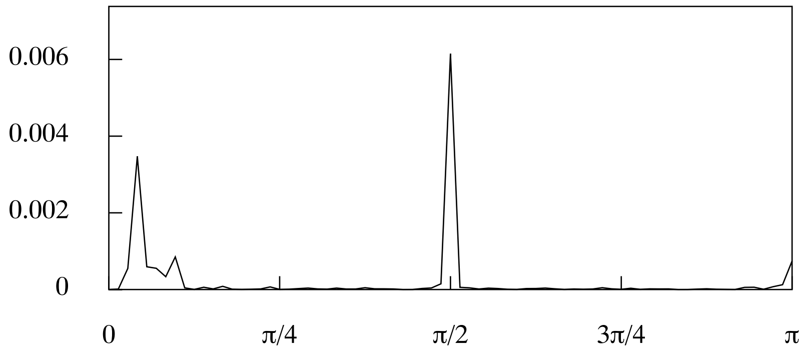

Figure 6: The periodogram of the residuals obtained by fitting a quadratic trend through the logarithmic sequence of U.K. income.

Example. Figure 4 displays a sequence of the logarithms of the quarterly series of U.K. Gross Domestic Product (GDP) over the period from 1970 to 2005. Interpolated through this sequence is a quadratic trend, which represents the growth path of the economy.

The deviations from this growth path are a combination of the low-frequency business cycle with the high-frequency fluctuations that are due to the seasonal nature of economic activity. These deviations are represented in Figure 5, which also shows an interpolated continuous function that is designed to represent the business cycle.

The periodogram of the deviations is shown in Figure 6. This gives a clear indication of the separability of the business cycle and the seasonal fluctuations. The spectral structure extending from zero frequency up to $\pi/8$ belongs to the business cycle. The prominent spikes located at the frequency $\pi/2$ and at the limiting Nyquist frequency of π are the property of the seasonal fluctuations. Elsewhere in the periodogram, there are wide dead spaces, which are punctuated by the spectral traces of minor elements of noise.

The slowly varying continuous function interpolated through the deviations of Figure 5 has been created by combining a set of sine and cosine functions of increasing frequencies in the manner of equation (27), but with the summation extending no further than the limiting frequency of the business cycle, which is $\pi/8$.

4.2 Filtering and the Frequency Domain

Given that a data sequence can be represented in terms of trigonometrical functions, it is appropriate to consider the effect of applying a linear filter to such elements. Mapping a (doubly-infinite) cosine sequence $y(t) = \cos(\omega t)$ through a filter defined by the coefficients $\{\psi_j\}$ produces the output

$$\begin{aligned} x(t) &= \sum_j \psi_j \cos(\omega[t-j]) \\ &= \sum_j \psi_j \cos(\omega j)\cos(\omega t) + \sum_j \psi_j \sin(\omega j)\sin(\omega t) \\ &= \alpha\cos(\omega t) + \beta\sin(\omega t) = \rho\cos(\omega t - \theta), \end{aligned} \tag{38}$$

where $\alpha = \sum_j \psi_j \cos(\omega j)$, $\beta = \sum_j \psi_j \sin(\omega j)$, $\rho = \surd(\alpha^2 + \beta^2)$ and $\theta = \tan^{-1}(\beta/\alpha)$. These results follow in view of the trigonometrical identity of (28).

The effect of the filter is to alter the amplitude of the cosine via the gain factor ρ and to induce a delay that corresponds to the phase angle θ. It is apparent that, if the filter is symmetric about the coefficient ψ_0, with $\psi_{-j} = \psi_j$, then $\beta = \sum_j \psi_j \sin(\omega j) = 0$ and, therefore, $\theta = 0$. That is to say, a symmetric filter that looks equally forward and backwards in time has no phase effect.

The z-transform of the sequence of filter coefficients is the polynomial

$$\psi(z) = \sum_j \psi_j z^j, \tag{39}$$

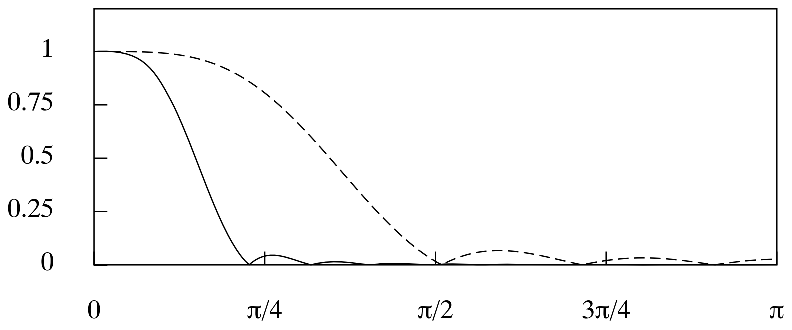

Figure 7: The gain functions of the Henderson filters of 9 coefficients (broken line) and 23 coefficients (continuous line).

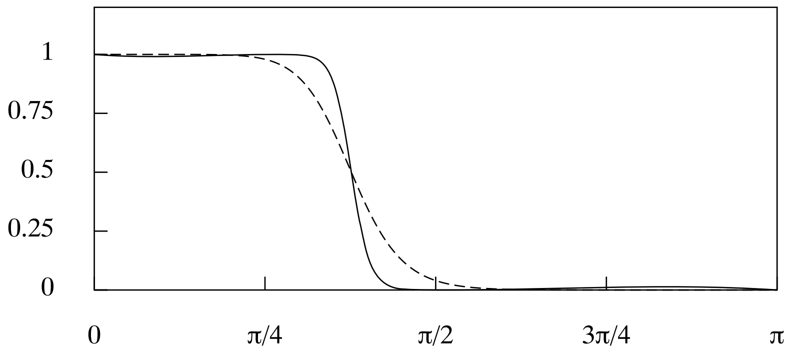

Figure 8: The gain functions of the Butterworth lowpass filters with $n = 4$ (broken line) and $n = 11$ (continuous line), both with a nominal cut-off frequency of $3\pi/8$ radians.

wherein z stands for a complex number. Setting $z = \exp\{-\mathrm{i}\omega\} = \cos(\omega) - \mathrm{i}\sin(\omega)$ constrains this number to lie on the unit circle in the complex plane. The resulting function

$$\begin{aligned} \psi(\exp\{-\mathrm{i}\omega\}) &= \sum_j \psi_j \cos(\omega j) - \mathrm{i}\sum_j \psi_j \sin(\omega j) \\ &= \alpha(\omega) - \mathrm{i}\beta(\omega) \end{aligned} \tag{40}$$

is the frequency response function, which is, in general, a periodic complex-valued function of ω with a period of 2π. In the case of a symmetric filter, it becomes a real-valued and even function, which is symmetric about $\omega = 0$. When the frequency response function is defined over the interval $[-\pi, \pi)$, or equally over the interval $[0, 2\pi)$, it conveys all of the information concerning the gain and the phase effects of the filter. For a more concise notation, we may write $\psi(\omega)$ in place of $\psi(\exp\{-\mathrm{i}\omega\})$.

An alternative expression for the frequency response function derives from the polar representation of complex numbers. We denote the squared modulus of the function $\psi(\omega) = \alpha(\omega) - \mathrm{i}\beta(\omega)$ by $|\psi(\omega)|^2 = \alpha^2(\omega) + \beta^2(\omega)$ and its argument by $\theta(\omega) = \tan^{-1}\{\beta(\omega)/\alpha(\omega)\}$. Then, there is

$$\begin{aligned} \psi(\omega) &= |\psi(\omega)|e^{-\mathrm{i}\theta(\omega)} \\ &= |\psi(\omega)|[\cos\{\theta(\omega)\} - \mathrm{i}\sin\{\theta(\omega)\}]. \end{aligned} \tag{41}$$

The function $|\psi(\omega)|$ describes the gain of the filter,

Example. Figure 7 represents the gain of the symmetric Henderson filters of $m = 9$ and $m = 23$ coefficients. The gain is unity at $\omega = 0$, which means that the filters preserve the trend component.

There is a gradual attenuation of the gain until it reaches zero, which is close to $\omega = \pi/2$ in the case of $m = 9$ and slightly below $\omega = \pi/4$ when $m = 23$. Thereafter, the gain fluctuates as the frequency increases. These fluctuations may be regarded as a design fault of the filter; and other designs may be sought that suppress the high-frequency components more firmly.

A filter design that has long been popular in electrical engineering, at least in its analogue form, is the Butterworth filter—see Pollock (2000). The digital form of the lowpass filter can be expressed in terms of the following rational function of z:

$$\psi(z) = \frac{(1+z)^n(1+z^{-1})^n}{(1+z)^n(1+z^{-1})^n + \lambda(1-z)^n(1-z^{-1})^n}. \tag{42}$$

The factors $(1+z)$ and $(1+z^{-1})$ in the numerator ensure that the gain is zero when $z = -1$, which is the case when $\omega = \pi$ within $z = \exp\{-\mathrm{i}\omega\}$. On the other hand, when $z = 1$, which is when $\omega = 0$, the factors $(1-z)$ and $(1-z^{-1})$ in the denominator are zeros; and the gain of the filter is unity. The mid point ω_c of the transition from unit gain to zero gain is governed by the so-called smoothing parameter $\lambda = \{1/\tan(\omega_c/2)\}^{2n}$; and the integer parameter n, described as the filter order, determines the rate of the transition between the two values.

Figure 8 shows that the Butterworth filters discriminate clearly between the pass band, where, ideally, the gain is unity, and the stop band, where the gain should be zero. In this respect, they are superior to the Henderson filters. However, since a Butterworth filter corresponds to a rational function of z, in contrast to the simple polynomial function of a Henderson filter, its implementation, which we shall describe in a later section, is less straightforward.

An implementation of the Butterworth filter is available on the compact disc that accompanies the book of Pollock (1999).

4.3 Aliasing and the Shannon–Nyquist Sampling Theorem

In equation (27), the frequencies of the trigonometric functions range from $\omega_1 = 2\pi/T$ to $\omega_n = \pi$. The frequency of π radians per sampling interval is the so-called Nyquist frequency. Although the process generating the data may contain components of frequencies higher than the Nyquist frequency, these will not be detected when it is sampled regularly at unit intervals of time. In fact, the effects on the process of components with frequencies in excess of the Nyquist value will be confounded with those whose frequencies fall below it.

To demonstrate this, consider the case where the process contains a component that is a pure cosine wave of unit amplitude and zero phase, whose frequency ω lies in the interval $\pi < \omega < 2\pi$. Let $\omega^* = 2\pi - \omega$. Then,

$$\begin{aligned} \cos(\omega t) &= \cos\bigl\{(2\pi - \omega^*)t\bigr\} \\ &= \cos(2\pi)\cos(\omega^* t) + \sin(2\pi)\sin(\omega^* t) \\ &= \cos(\omega^* t); \end{aligned} \tag{43}$$

which indicates that ω and ω^* are observationally indistinguishable. Here, $\omega^* < \pi$ is described as the alias of $\omega > \pi$.

We have demonstrated that, if the rate of sampling is too low, then the resulting sample misrepresents the underlying process. We also need to show that, if the sampling rate is sufficient, then a continuous function can be represented without loss of information by a discrete sample. Thus, according to the Shannon–Nyquist sampling theorem, any square integrable continuous function $x(t)$ that has a Fourier transform $\xi(\omega)$ that is band-limited in the frequency domain, with $\xi(\omega) = 0$ for $\omega > \pi$, has the series expansion

$$x(t) = \sum_{k=-\infty}^{\infty} x_k \frac{\sin\{\pi(t-k)\}}{\pi(t-k)} = \sum_{k=-\infty}^{\infty} x_k \psi(t-k), \tag{44}$$

where $x_k = x(k)$ is the value of the function $x(t)$ at the point $t = k$. It follow that the continuous function $x(t)$ can be reconstituted from its sampled values $\{x_t; t \in \mathcal{I}\}$. Observe that $\psi(t-k)$ is just a displaced version of the sinc function illustrated in Figure 3.

In proving this, we use the result that, if $x(t)$ is a continuous square-integrable function, then it is amenable to a Fourier integral transform, which gives

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \xi(\omega)e^{i\omega t}d\omega, \quad \text{where} \quad \xi(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt. \tag{45}$$

But $\xi(\omega)$ is a continuous band-limited function defined on the interval $(-\pi, \pi]$ that may also be regarded as a periodic function of a period of 2π. Therefore, $\xi(\omega)$ is amenable to a classical Fourier analysis; and it may be expanded as

$$\xi(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{-ik\omega}, \quad \text{where} \quad c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \xi(\omega)e^{ik\omega}d\omega. \tag{46}$$

By comparing (45) with (46), we see that the coefficients c_k are simply the ordinates of the function $x(t)$ sampled at the integer points; and we may write them as

$$c_k = x_k = x(k). \tag{47}$$

Next, we must show how the continuous function $x(t)$ may be reconstituted from its sampled values. Using (47) in (46) gives

$$\xi(\omega) = \sum_{k=-\infty}^{\infty} x_k e^{-ik\omega}. \tag{48}$$

Putting this in (45), and taking the integral over $[-\pi, \pi]$ in consequence of the band-limited nature of the function $x(t)$, gives

$$x(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\sum_{k=-\infty}^{\infty} x_k e^{-ik\omega}\right\}e^{i\omega t}d\omega = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} x_k \int_{-\pi}^{\pi} e^{i\omega(t-k)}d\omega. \tag{49}$$

The integral on the RHS is evaluated as

$$\int_{-\pi}^{\pi} e^{i\omega(t-k)}d\omega = 2\frac{\sin\{\pi(t-k)\}}{t-k}. \tag{50}$$

Putting this into the RHS of (49) gives the result of (44).

The Shannon–Nyquist theorem concerns the representation of a square integrable function in terms a sequence of sinc functions weighted by coefficients that constitute a square summable sequence. However, we shall also be concerned with the representation of a continuous stationary stochastic process defined over the real line and band-limited in frequency to the interval $[-\pi, \pi]$. Such a function is not integrable over the entire real line.

Nevertheless, the Shannon–Nyquist theory can generalised to accommodate this difficulty; and the doubly-infinite set of sinc functions $\{\sin(\pi[t-k])/(t-k); k = 0 \pm 1, \pm 2, \ldots\}$ provides a basis for all such continuous functions, which is, in fact, an orthogonal basis.

In practice, we deal only with finite data sequences; and, for the purposes of Fourier analysis, the data can be wrapped around a circle of a circumference T, equal to the number of data points. This is tantamount to regarding the data sequence as one cycle of a periodic sequence.

The periodic kernel function that would be used for interpolating a continuous periodic function through these data points is the Dirichlet kernel. This can be derived from the sinc function, which has an infinite support, by wrapping it around the circle and by adding the overlying layers. The process of interpolation based on the Dirichlet kernel corresponds exactly with the process of Fourier synthesis which is based on the spectral ordinates of the data.

The sampling theorem has been attributed to several authors, including Whittaker (1915) who published the interpolation formula. The origins of the sampling theorem have been described by Luke (1999), and Higgins (1985) has described the development of the interpolation formula.

4.4 The Processes Underlying the Data

It is appropriate to consider, in the light of the phenomenon of aliasing and of the –Nyquist sampling theorem, the nature of the processes underlying the discretely sampled data. A stationary data sequence is commonly modelled as a ARMA process, which is the result of applying a rational transfer function or filter to a white-noise sequence of independently and identically distributed random variables. However, the provenance of the white-noise sequence itself requires some explanation.

A common explanation is that the white noise originates in the sampling of a continuous-time Wiener process. The latter is the product of the cumulation of a stream of infinitesimal impulses, which are the stationary and independent increments of the process.

An impulse in continuous time has a uniform power spectrum that is distributed over the entire frequency range, with infinitesimal power in any finite interval. The sampling of a Wiener process at regular intervals entails a process of aliasing whereby the cumulated increments gives rise to a uniform spectrum of finite power over the interval $[-\pi, \pi]$. The advantage of this conceptualisation is that it places the origin of the white-noise sequence is an identifiable continuous-time process.

An alternative approach is to imagine that the discrete-time white noise is derived by sampling a continuous-time process that is inherently limited in frequency to the interval $[0, \pi]$. Figure 3 depicts a continuous sinc function of which the Fourier transform is a rectangle on the frequency interval $[-\pi, \pi]$. The sinc function can also be construed as a wave packet centred on time $t = 0$. A continuous stochastic function that is band-limited by the Nyquist frequency of π will be generated by the superposition of such functions arriving at regular or irregular intervals of time and having amplitudes that are randomly distributed.

(The wave packet is a concept from quantum mechanics—see, for example Dirac (1958) and Schiff (1981)—that has become familiar to statisticians via wavelet analysis—see, for example, Daubechies (2004) and Percival and Walden (2000).)

A continuous-time counterpart of a discrete-time white-noise process, from which the latter may be obtained by sampling, can be constituted from a stream of sinc function wave packets arriving at unit time intervals and having amplitudes that are independently and identically distributed with a zero mean and a common variance. This is indeed an artificial process, but it is viable in theory and it has the advantage of making no reference to the phenomenon of aliasing.

A further advantage of this concept of white noise is that it permits us to define a band-limited white noise that has an upper frequency bound that is less than the Nyquist frequency of π or a lower frequency bound that is greater than zero, or one that has both of these features. A sinc function wave packet that is limited, in positive frequencies, to the band $[\alpha, \beta] \in [0, \pi]$, which would be the basis of white noise confined to that band, has the functional form of

$$\begin{aligned} \psi(t) &= \frac{1}{\pi t}\{\sin(\beta t) - \sin(\alpha t)\} \qquad (51) \\ &= \frac{2}{\pi t}\cos\{(\alpha+\beta)t/2\}\sin\{(\beta-\alpha)t/2\} \\ &= \frac{2}{\pi t}\cos(\gamma t)\sin(\delta t), \end{aligned}$$

where $\gamma = (\alpha + \beta)/2$ is the centre of the band and $\delta = (\beta - \alpha)/2$ is half its width. The equality follows from the identity $\sin(A + B) - \sin(A - B) = 2\cos A \sin B$.

It has been shown by Pollock and Lo Cascio (2006) that, when the interval $[0, \pi]$ is partitioned by a sequence of p frequency bands of equal width, an orthogonal basis can be obtained for each band by displacing its wavelets successively by p elements at a time. The implication is that, to obtain full information on a process that is limited to such a band, we need only sample it at the rate of one observation in p sample periods.

In the case of the business cycle function of Figure 5, which is band-limited to the frequency interval $[0, \pi/8]$, only one in eight of the points sampled from this function at unit time intervals needs to be retained in order to convey all of the relevant information. The periodogram of the resulting subsampled sequence, which has the frequency range of $[0, \pi]$, will have a profile identical to that of the spectral structure that occupies the interval $[0, \pi/8]$ in Figure 6.

A conventional ARMA model depicts a stationary process that has a spectral density function that is nonzero everywhere in the interval $[0, \pi]$, except, possibly, over a set of measure zero. Such models are inappropriate to the business cycle data of Figure 5, which have a periodogram that is effectively zero-valued over large intervals of the frequency range. However, there should be no difficulty in fitting an ARMA model to data obtained by subsampling the business-cycle sequence that underlies the smooth trajectory of Figure 5 at the rate of one in eight.

5 The Classical Wiener–Kolmogorov Theory

The purpose of a Wiener–Kolmogorov filter is to extract an estimate of a signal sequence $\xi(t)$ from an observable data sequence

$$y(t) = \xi(t) + \eta(t), \tag{52}$$

which is afflicted by the noise $\eta(t)$. The theory was formulated independently by Norbert Wiener (1941) and Andrei Nikolaevich Kolmogorov (1941) during the Second World War. They were both considering the problem of how to target radar-assisted anti-aircraft guns on incoming enemy aircraft.

According to the classical assumptions, which we shall later amend, the signal and the noise are generated by zero-mean stationary stochastic processes that are mutually independent. It follows that the autocovariance generating function of the data is the sum of the autocovariance generating functions of its two components. Thus

$$\gamma^{yy}(z) = \gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z) \quad \text{and} \quad \gamma^{\xi\xi}(z) = \gamma^{y\xi}(z). \tag{53}$$

These functions are amenable to the so-called Cramér–Wold factorisation, and they may be written as

$$\gamma^{yy}(z) = \phi(z^{-1})\phi(z), \quad \gamma^{\xi\xi}(z) = \theta(z^{-1})\theta(z), \quad \gamma^{\eta\eta}(z) = \theta_\eta(z^{-1})\theta_\eta(z). \tag{54}$$

Such factorisations were considered by Wold (1954), who cited unpublished work of H. Cramér. An effective algorithm for achieving the factorisation was proposed by Tunnicliffe–Wilson (Wilson, 1969) and the code for implementing it has been provided by Pollock (1999), amongst others. For a further discussion, see Godolphin (1976).

The estimate of the signal element ξ_t is a linear combination

$$x_t = \sum_{j=-p}^{q} \psi_{t,j} y_{t-j} \tag{55}$$

of the available data points within the information vector

$$y = [y_{t-q}, y_{t+1-q}, \ldots, y_t, \ldots, y_{t+p-1}, y_{t+p}]'.$$

The vector may contain all of the available data, or it may represent a narrow window that is moving over the data. If the contents of the data window are fixed and if the window does not move forward with each successive estimate of the signal, then the coefficients $\psi_{t,j}$ of the filter are liable to vary with both t and j.

The principle of minimum-mean-square-error estimation indicates that the estimation errors must be statistically uncorrelated with the elements of the information set. Thus

$$\begin{aligned} 0 &= E\big\{y_{t-j}(\xi_t - x_t)\big\} \qquad (56) \\ &= E(y_{t-j}\xi_t) - \sum_{k=-p}^{q} \psi_{t,k} E(y_{t-j}y_{t-k}) \\ &= \gamma_j^{y\xi} - \sum_{k=-p}^{q} \psi_{t,k} \gamma_{j-k}^{yy}. \end{aligned}$$

Equation (56) can be rendered also in a matrix format. By running from $j = q$ to $j = -p$, we get the following system:

$$\begin{bmatrix} \gamma_q^{\xi\xi} \\ \gamma_{q-1}^{\xi\xi} \\ \vdots \\ \gamma_p^{\xi\xi} \end{bmatrix} = \begin{bmatrix} \gamma_0^{yy} & \gamma_1^{yy} & \cdots & \gamma_{p+q}^{yy} \\ \gamma_1^{yy} & \gamma_0^{yy} & \cdots & \gamma_{p+q-1}^{yy} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p+q}^{yy} & \gamma_{p+q-1}^{yy} & \cdots & \gamma_0^{yy} \end{bmatrix} \begin{bmatrix} \psi_{t,q} \\ \psi_{t,q-1} \\ \vdots \\ \psi_{t,-p} \end{bmatrix}. \tag{57}$$

Here, on the LHS, we have set $\gamma_j^{y\xi} = \gamma_j^{\xi\xi}$ in accordance with (53).

Let $T = p + q + 1$ be the number of elements in the information set, and define the dispersion matrices Ω_ξ, Ω_η and $\Omega_y = \Omega_\xi + \Omega_\eta$ of order T of the vectors ξ, η and $y = \xi + \eta$, which contain the elements of the signal, the noise and the data that fall within the span of the information set. Then, equation (57) can be written in summary notation as $\Omega_\xi e_q = \Omega_y \psi_{t\bullet}'$, where e_q is a vector of order T containing a single unit preceded by q zeros and followed by p zeros. The coefficient vector $\psi_{t\bullet} = [\psi_{t,q}, \psi_{t,q-1}, \ldots, \psi_{t,-p}]$ is given by

$$\psi_{t\bullet} = e_q'\Omega_\xi\Omega_y^{-1} = e_q'\Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}. \tag{58}$$

and the estimate of ξ_t is $x_t = \psi_{t\bullet}y$. (There are p data elements ahead of this prediction for time t and q behind, which accounts for e_q.) Given $y = [y_0, y_1 \ldots, y_{T-1}]'$, which contains all of the available data, the estimate of the complete vector $\xi = [\xi_0, \xi_1 \ldots, \xi_{T-1}]'$ of the corresponding signal elements would be

$$x = \Omega_\xi\Omega_y^{-1}y = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y. \tag{59}$$

This is the finite-sample version of the Wiener–Kolmogorov filter, which will be discussed more fully in section 7.

Observe that equation (59) represents a time-varying filter. The tth row of the matrix $\Omega_\xi\Omega_y^{-1}$ provides the filter coefficients that serve to generate the value x_t by a combination of the elements of the data vector y. It also worth noting that this filter requires no extrapolations of the data to enable it to reach the ends of the sample.

The classical Wiener–Kolmogorov theory was aimed at developing linear time-invariant filters that would be appropriate to semi-infinite and doubly infinite data sequences; and many of the subsequent developments have taken the classical results as their starting point.

To derive the classical formulae, consider suppressing the time subscript of $\psi_{t,k}$ within equation (56). On multiplying throughout by z^j, the equation can be rendered as

$$\begin{aligned} \gamma_j^{y\xi}z^j &= z^j(\psi_{-p}\gamma_{j+p}^{yy} + \psi_{1-p}\gamma_{j-1+p}^{yy} + \cdots + \psi_q\gamma_{j-q}^{yy}) \\ &= (\psi_{-p}z^{-p})(\gamma_{j+p}^{yy}z^{j+p}) + (\psi_{1-p}z^{1-p})(\gamma_{j-1-p}^{yy}z^{j-1-p}) + \\ &\qquad \cdots + (\psi_q z^q)(\gamma_{j-q}^{yy}z^{j-q}); \end{aligned} \tag{60}$$

and, in the case where $-p \leq j, k \leq q$, the full set of so-called normal equations can be expressed as

$$\gamma^{\xi\xi}(z)_{(-p,q)} = \left[\gamma^{yy}(z)\psi(z)\right]_{(-p,q)}, \tag{61}$$

where we have set $\gamma^{y\xi}(z) = \gamma^{\xi\xi}(z)$, according to (53), and where the subscript $(-p, q)$ indicates that only the terms associated with $z^{-p}, z^{1-p}, \ldots, z^q$ have been taken from $\gamma^{\xi\xi}(z)$ and $\gamma^{yy}(z)\psi(z)$.

Equation (61) can accommodate a wide variety of assumptions concerning the extent of the information set. These include the case of a causal FIR (finite impulse response) filter (with $j \in [0, n]$), a symmetric two-sided FIR filter (with $j \in [-n, n]$), a causal IIR (infinite impulse response) filter (with $j \in [0, \infty]$), or a bidirectional IIR filter (with no bounds on j).

In the case of a causal IIR filter, the normal equations take the form of

$$\left[\gamma^{\xi\xi}(z)\right]_+ = \left[\phi(z^{-1})\phi(z)\psi(z)\right]_+, \tag{62}$$

where the subscripted + is to indicate that only the part of the series which contains nonnegative powers of z is to be taken. (This is the notation of Whittle (1983).) The solution is

$$\psi(z) = \frac{1}{\phi(z)}\left[\frac{\gamma^{\xi\xi}(z)}{\phi(z^{-1})}\right]_+. \tag{63}$$

The classical Wiener–Kolmogorov theory also considers the case of a doubly infinite information set. When there are no restrictions on the exponent of z, the normal equations of (61) become $\gamma^{\xi\xi}(z) = \gamma^{yy}(z)\psi(z)$. Then, there is

$$\psi(z) = \frac{\gamma^{\xi\xi}(z)}{\gamma^{yy}(z)} = \frac{\theta(z^{-1})\theta(z)}{\phi(z^{-1})\phi(z)}, \tag{64}$$

which is the basis for a symmetric IIR filter. An example is provided by the Butterworth filter of (42).

Notwithstanding the fact that equation (57) provides an appropriate environment in which to derive filters for finite data sequences, it has been customary to derive such filters in reference to equation (64), which relates to data sequences that are doubly infinite. A filter that presupposes an infinite data set is unrealisable in practice; and there are three common ways of deriving a practical filter from equation (64).

The first way, which is the simplest, depends upon generating the Laurent expansion of the rational function to create the series $\psi(z) = \{\psi_0 + \psi_1(z^{-1} + z) + \cdots\}$. From the central coefficients of the expansion, an FIR filter is formed, which can be applied to the data. However, to obtain a good approximation to the theoretical filter of (64), a large number of coefficients may be needed. Hillmer and Tiao (1982) have used such a method.

The second method of implementing the filter depends upon the Cramér–Wold factorisations of $\gamma^{\xi\xi}(z)$ and $\gamma^{yy}(z)$. From the resulting factors, two filters can be formed, one working in direct time and the other in reverse time. The filtering operations may be represented by

$$\phi(z)q(z) = \theta(z)y(z), \qquad \phi(z^{-1})x(z) = \theta(z^{-1})q(z). \tag{65}$$

The first filter, which runs forwards in time, generates the intermediate output $q(t)$, and the second filter, which runs backwards in time, generates the final output $x(t)$. This is the bidirectional method, which was been the leitmotif of the method proposed by Pollock (2000), which amounts to a procedure for solving equation (59).

In the third method, a factorisation is employed that has the form of

$$\frac{\gamma^{\xi\xi}(z)}{\gamma^{yy}(z)} = \frac{\rho(z)}{\phi(z)} + \frac{\rho(z^{-1})}{\phi(z^{-1})}. \tag{66}$$

Two parallel sequences $f(t)$ and $b(t)$ are generated via

$$\rho(z)f(z) = \theta(z)y(z), \qquad \rho(z^{-1})b(z) = \theta(z^{-1})y(z), \tag{67}$$

and the results are added to create $x(t) = f(t) + b(t)$. This is the contragrade method of Burman (1980), who attributed it to G. Tunnicliffe–Wilson. It has been employed in the TRAM0–SEATS program of Gómez and Maravall (1996) and of Caporello and Maravall (2004).

In the econometric analysis of time series, it is common to model a nonstationary process via an autoregressive operator with roots of unit value, which are on the boundary of instability. Imagine that the data sequence $y(t) = \xi(t) + \eta(t)$ contains a stationary noise component $\eta(t)$ and a nonstationary trend component $\xi(t)$ that can be reduced to stationarity by p-fold differencing. Let $\nabla(z) = 1 - z$ be the z-transform of the difference operator. Then, multiplying $y(z)$ by $\nabla^p(z)$ will give

$$\begin{aligned} \nabla^p(z)y(z) &= \nabla^p(z)\xi(z) + \nabla^p(z)\eta(z) \\ &= \delta(z) + \kappa(z) = g(z). \end{aligned} \tag{68}$$

The estimates of the differenced components $\delta(t)$ and $\kappa(t)$ may be denoted by $d(t)$ and $k(t)$ respectively. They may be extracted from the differenced data $g(t)$ in the various ways that have already been described for stationary data, using filters based on

$$\psi_\delta(z) = \frac{\gamma^{\delta\delta}(z)}{\gamma^{gg}(z)} = \frac{\gamma^{\xi\xi}(z)}{\gamma^{yy}(z)} \quad \text{and} \quad \psi_\kappa(z) = \frac{\gamma^{\kappa\kappa}(z)}{\gamma^{gg}(z)} = \frac{\gamma^{\eta\eta}(z)}{\gamma^{yy}(z)}, \tag{69}$$

which are complementary in the sense that $\psi_\delta(z) + \psi_\kappa(z) = 1$. Thereafter, the sought-after estimates of $\xi(t)$ and $\eta(t)$, denoted by $x(t)$ and $h(t)$, respectively, can be obtained by cumulating their differenced versions.

Let $\Sigma(z) = \nabla^{-1}(z)$ denote the cumulation operator, which is the inverse of the difference operator. Then, $h(z) = \Sigma^p(z)k(z)$ is the z-transform of the sequence $h(t)$, obtained by cumulating $k(t)$. The latter is given by the $k(z) = \psi_\kappa(z)g(z)$. When $h(t)$ is available, the estimate of $\xi(t)$ may be obtained by subtraction:

$$x(t) = y(t) - h(t). \tag{70}$$

If $\psi_\kappa(z)$ contains the factor $(1-z)^n$ of a degree $n \geq p$—which will often prove to be the case—then applying the reduced filter $\psi_\kappa^*(z) = (1-z)^{-p}\psi_\kappa(z) = \Sigma^p(z)\psi_\kappa(z)$ to $g(z)$ will produce $h(z)$ directly. Thus, one can avoid the need to cumulate the filtered sequence, which means that there will be no need for starting values. Observe also that $\psi_\kappa^*(z)g(z) = \psi_\kappa(z)y(z)$, so we might apply the original filter to the undifferenced data. However, this would require us to supply nonzero starting values to the filter.

In the next section, we shall consider in more detail the means of converting to a matrix format a set of equations that have been expressed in terms of the z-transform; and, thereafter, we shall be considering the estimation of the time-varying filter coefficients in more detail.

6 Matrix Formulations

The classical theory of statistical signal extraction presupposes lengthy data sequences, which are assumed, in theory, to be doubly infinite or semi-infinite, and it is also assumed that the processes generating these data are statistically stationary. In many practical cases, and in most econometric applications, the available data are, to the contrary, both strongly trended and of a limited duration.

In order to adapt the theory to these circumstances, it is helpful to employ a formulation that is in terms of matrices and vectors. The theory of local polynomial regression, which has been touched on in the section 3, and which entails FIR filters, is naturally expressed in matrices. The filtering theory that has been developed by electrical engineers, and which typically makes reference to the frequency domain, has been expressed in terms of the z-transforms of the data sequences and of the sequences of filter coefficients, which thereby become polynomial operators. These notational and conceptual differences have created a schism in the theory of statistical signal extraction that needs to be overcome.

A fruitful approach to unifying the theory is to seek a matrix representation of the argument z of the z-transforms. The latter is commonly interpreted as an arbitrary point in the complex plane, when it is not constrained to lie on the unit circle. Within the time domain, the argument assumes the role of a temporal lag operator or delay operator. In fact, in electrical engineering, the delay operator is commonly represented by z^{-1}; but it serves our present purposes better to depart from this convention by using z instead. There are two alternative matrix representations of the argument that preserve the underlying algebra of polynomials and rational functions to differing degrees.

6.1 Toeplitz Matrices

In the first of these matrix representations, which is appropriate to a time-domain interpretation of linear filtering, the argument z is replaced by

$$L_T = [e_1, e_2, \ldots, e_{T-1}, 0], \tag{71}$$

which is obtained 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. This is the finite-sample version of the lag operator L that is commonly employed by econometricians. When it is applied to the sequence $x(t) = \{x_t; t = 0, \pm1, \pm2, \ldots\}$, the effect of the lag operator is that $Lx(t) = x(t-1)$. The inverse element z^{-1}, which stands for the foreward-shift operator, can be replaced by the matrix

$$F_T = [0, e_1, e_2, \ldots, e_{T-2}] = L_T'. \tag{72}$$

We note that, whereas $zz^{-1} = z^{-1}z = 1$, which is an identity operator, there is $L_TF_T \neq F_TL_T \neq I_T$. Here, the discrepancy lies in the fact that L_TF_T differs from I_T in having a zero as its leading, top-left, element whereas F_TL_T differs from I_T in having a zero for its final, bottom-right, element.

A related discrepancy is that, whereas z can be raised to any power, the operators L_T and F_T are nilpotent of degree T, which is to say that $L_T^{T+q} = 0$ and $F_T^{T+q} = 0$ for all $q \geq 0$.

Given a Laurent polynomial of the form $\alpha(z) = \sum_{j=-p}^{q} \alpha_j z^j$, we can replace the powers of z and z^{-1} by powers of L_T and F_T, respectively, to obtain the banded Toeplitz matrix $A = [\alpha_{i-j}]$, in which the generic element in the ith row and the jth column is $\alpha_{ij} = \alpha_{i-j}$. More particularly, if $\alpha(z) = \sum_{j=0}^{q} \alpha_j z^j$ is a polynomial in positive powers of z, then replacing z by L_T gives rise to a lower-triangular Toeplitz matrix $A = \alpha(L_T)$, whereas replacing z^{-1} in $\alpha(z^{-1}) = \sum_{j=0}^{q} \alpha_j z^{-j}$ by F_T leads to the corresponding upper-triangular matrix. A lower-triangular Toeplitz matrix $A = \alpha(L_T)$ is completely characterised by its leading column, which is $\alpha = Ae_0$ whereas the upper-triangular matrix $A' = \alpha(F_T)$ is completely characterised by is leading row $\alpha' = e_0'A'$.

It is important to note that lower-triangular (LT) Toeplitz matrices commute in multiplication as do the upper triangular matrices. This is attributable to their origins in polynomials. Thus

$$\begin{gathered}\text{If } A = \alpha(L_T) \text{ and } B = \beta(L_T) \text{ are LT Toeplitz matrices,}\\ \text{then } AB = BA \text{ is also an LT Toeplitz matrix.}\end{gathered} \tag{73}$$

It also follows that

$$ABe_0 = A\beta = BAe_0 = B\alpha. \tag{74}$$

It is of some significance that this commutativity in multiplication does not extend to Toeplitz matrices in general. If fact, if $A = [\alpha_{i-j}]$ and $B = [\beta_{i-j}]$ are Toeplitz matrices, then $AB = (BA)^{\#}$ and $BA = (AB)^{\#}$, where $Q^{\#}$ denotes the counter transpose of Q, which is its reflection about the secondary SW–NE diagonal.

Example. Consider a sequence $y(t) = \{y_t; t = 0, \pm 1, \pm 2, \ldots\}$ generated by a moving-average process of order q. Then,

$$y_t = \sum_{j=0}^{q} \mu_j \varepsilon_{t-j} \quad \text{with} \quad \mu_0 = 1, \tag{75}$$

where ε_t is from a white-noise sequence of independently and identically distributed random variables of a zero mean and a finite variance σ_ε^2. If $\mu(z) = \sum_{j=0}^{q} \mu_j z^j$ is the z-transform of the moving-average coefficients, then $\gamma(z) = \sigma_\varepsilon^2 \mu(z^{-1})\mu(z)$ is the autocovariance generating function of the process.

Now consider a vector $y = [y_0, y_1, \ldots, y_{T-1}]'$ of T observations sampled from the process. This can be written as

$$y = M_* \varepsilon_* + M\varepsilon, \tag{76}$$

where $\varepsilon = [\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{T-1}]'$contains disturbances from within the sample period and $\varepsilon_* = [\varepsilon_{-q}, \ldots, \varepsilon_{-2}, \varepsilon_{-1}]'$ is a vector of presample elements.

The matrix $M = \mu(L_T)$, which is of a lower-triangular Toeplitz form, is completely characterised by its leading vector $[\mu_0, \mu_1, \ldots, \mu_q, 0, \ldots 0]'$. The matrix $M_* = [M_{**}', 0]'$ contains the parameters associated with the presample elements.

An example is provided by the following display that relates to the case where the moving-average order is $q = 3$ and the size of the sample is $T = 6$:

$$M_* = \begin{bmatrix} \mu_3 & \mu_2 & \mu_1 \\ 0 & \mu_3 & \mu_2 \\ 0 & 0 & \mu_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad M = \begin{bmatrix} \mu_0 & 0 & 0 & 0 & 0 & 0 \\ \mu_1 & \mu_0 & 0 & 0 & 0 & 0 \\ \mu_2 & \mu_1 & \mu_0 & 0 & 0 & 0 \\ \mu_3 & \mu_2 & \mu_1 & \mu_0 & 0 & 0 \\ 0 & \mu_3 & \mu_2 & \mu_1 & \mu_0 & 0 \\ 0 & 0 & \mu_3 & \mu_2 & \mu_1 & \mu_0 \end{bmatrix}. \tag{77}$$

The autocovariance matrix of the sample is given by

$$\begin{aligned}\Omega &= \gamma_0 I_T + \sum_{j=1}^{q} \gamma_j (L_T^j + F_T^j) \\ &= \sigma_\varepsilon^2 (M_* M_*' + MM').\end{aligned} \tag{78}$$

The first expression on the RHS is obtained directly from the autocovariance generating function by replacing the powers of z and z^{-1} by the corresponding powers of L_T and F_T and by replacing $z^0 = 1$ by I_T. The second expression comes from (76), when the latter is used within $E(yy') = \Omega$.

It is notable that, whereas $\gamma(z) = \sigma_\varepsilon^2 \mu(z)\mu(z^{-1}) = \sigma_\varepsilon^2 \mu(z^{-1})\mu(z)$, we find that $\Omega \neq \sigma_\varepsilon^2 MM' \neq \sigma_\varepsilon^2 M'M$, where M and Ω are the direct matrix analogues of $\mu(z)$ and $\gamma(z)$. The difficulty, which resides in the leading submatrix of MM' and the trailing submatrix of $M'M$ is an end-of-sample problem due to the extra-sample elements.

For a discussion of Toeplitz matrices in relation to the maximum-likelihood estimator of a Gaussian autoregressive-moving average process, see Godolphin and Unwin (1983).

6.2 Circulant Matrices

In the second of the matrix representations, which is appropriate to a frequency-domain interpretation of filtering, the argument z is replaced by the full-rank circulant matrix

$$K_T = [e_1, e_2, \ldots, e_{T-1}, e_0], \tag{79}$$

which is obtained from the identity matrix $I_T = [e_0, e_1, \ldots, e_{T-1}]$ by displacing the leading column to the end of the array. This is an orthonormal matrix of which the transpose is the inverse, such that $K_T' K_T = K_T K_T' = I_T$. The powers of the matrix form a T-periodic sequence such that $K_T^{T+q} = K_T^q$ for all q. The periodicity of these powers is analogous to the periodicity of the powers of the argument $z = \exp\{-\mathrm{i}2\pi/T\}$, which is to be found in the Fourier transform of a sequence of order T.

The matrices $K_T^0 = I_T, K_T, \ldots, K_T^{T-1}$ form a basis for the set of all circulant matrices of order T—a circulant matrix $X = [x_{ij}]$ of order T being defined as a matrix in which the value of the generic element x_{ij} is determined by the index $\{(i-j) \bmod T\}$. This implies that each column of X is equal to the previous column rotated downwards by one element.

It follows that there exists a one-to-one correspondence between the set of all polynomials of degree less than T and the set of all circulant matrices of order T. Therefore, if $\alpha(z)$ is a polynomial of degree less that T, then there exits a corresponding circulant matrix

$$A = \alpha(K_T) = \alpha_0 I_T + \alpha_1 K_T + \cdots + \alpha_{T-1} K_T^{T-1}. \tag{80}$$

A convergent sequence of an indefinite length can also be mapped into a circulant matrix. Thus, if $\{\gamma_i\}$ is an absolutely summable sequence obeying the condition that $\sum |\gamma_i| < \infty$, then the z-transform of the sequence, which is defined by $\gamma(z) = \sum \gamma_j z^j$, is an analytic function on the unit circle. In that case, replacing z by K_T gives rise to a circulant matrix $\Gamma = \gamma(K_T)$ with finite-valued elements. In consequence of the periodicity of the powers of K_T, it follows that

$$\begin{aligned}\Gamma &= \left\{\sum_{j=0}^{\infty} \gamma_{jT}\right\} I_T + \left\{\sum_{j=0}^{\infty} \gamma_{(jT+1)}\right\} K_T + \cdots + \left\{\sum_{j=0}^{\infty} \gamma_{(jT+T-1)}\right\} K_T^{T-1} \\ &= \varphi_0 I_T + \varphi_1 K_T + \cdots + \varphi_{T-1} K_T^{T-1}.\end{aligned} \tag{81}$$

Given that $\{\gamma_i\}$ is a convergent sequence, it follows that the sequence of the matrix coefficients $\{\varphi_0, \varphi_1, \ldots, \varphi_{T-1}\}$ converges to $\{\gamma_0, \gamma_1, \ldots, \gamma_{T-1}\}$ as T increases. Notice that the matrix $\varphi(K) = \varphi_0 I_T + \varphi_1 K_T + \cdots + \varphi_{T-1} K_T^{T-1}$, which is derived from a polynomial $\varphi(z)$ of degree $T-1$, is a synonym for the matrix $\gamma(K_T)$, which is derived from the z-transform of an infinite convergent sequence.

The polynomial representation is enough to establish that circulant matrices commute in multiplication and that their product is also a polynomial in K_T. That is to say

$$\begin{aligned}&\text{If } X = x(K_T) \text{ and } Y = y(K_T) \text{ are circulant matrices,}\\ &\text{then } XY = YX \text{ is also a circulant matrix.}\end{aligned} \tag{82}$$

The matrix operator K_T has a spectral factorisation that is particularly useful in analysing the properties of the discrete Fourier transform. To demonstrate this factorisation, we must first define the so-called Fourier matrix. This is a symmetric matrix

$$U_T = T^{-1/2}[W_T^{jt}; t, j = 0, \ldots, T-1], \tag{83}$$

of which the generic element in the jth row and tth column is

$$\begin{aligned} W_T^{jt} \quad &= \quad \exp(-\mathrm{i}2\pi tj/T) = \cos(\omega_j t) - \mathrm{i}\sin(\omega_j t), \\ &\qquad \text{where} \quad \omega_j = 2\pi j/T. \end{aligned} \tag{84}$$

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

$$\bar{U}_T U_T = U_T \bar{U}_T = I_T, \tag{85}$$

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

The operator can be factorised as

$$K_T = \bar{U}_T D_T U_T = U_T \bar{D}_T \bar{U}_T, \tag{86}$$

where

$$D_T = \mathrm{diag}\{1, W, W^2, \ldots, W^{T-1}\} \tag{87}$$

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_T is T-periodic, such that $D_T^{q+T} = D_T^q$, and that $K_T^q = \bar{U}_T D_T^q U_T = U_T \bar{D}_T^q \bar{U}_T$ for any integer q. Since the powers of K_T form the basis for the set of circulant matrices, it follows that such matrices are amenable to a spectral factorisation based on (86).

Example. Consider, in particular, the circulant autocovariance matrix that is obtained by replacing the argument z in the autocovariance generating function $\gamma(z)$ by the matrix K_T. Imagine that the autocovariances form a doubly infinite sequence, as is the case for an autoregressive process or an autoregressive moving-average process:

$$\begin{aligned} \Omega^\circ = \gamma(K_T) \quad &= \quad \gamma_0 I_T + \sum_{\tau=1}^{\infty} \gamma_\tau (K_T^\tau + K_T^{-\tau}) \\ &= \quad \varphi_0 I_T + \sum_{\tau=1}^{T-1} \varphi_\tau (K_T^\tau + K_T^{-\tau}). \end{aligned} \tag{88}$$

Here, $\varphi_\tau; \tau = 0, \ldots, T-1$ are the "wrapped" coefficients that are obtained from the original coefficients of the autocovariance generating function in the manner indicated by (81). The spectral factorisation gives

$$\Omega^\circ = \gamma(K_T) = \bar{U}\gamma(D)U. \tag{89}$$

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

$$\gamma(\exp\{\mathrm{i}\omega_j\}) = \gamma_0 + 2\sum_{\tau=1}^{\infty} \gamma_\tau \cos(\omega_j \tau). \tag{90}$$

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 ω_j from the spectral density function of the linear (i.e. non-circular) stationary stochastic process.

7 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 chapter, 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

$$y = \xi + \eta, \tag{91}$$

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.

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 of (71). The resulting first and second order moments are denoted by

$$\begin{aligned} E(\xi) = 0, &\qquad D(\xi) = \Omega_\xi, \\ E(\eta) = 0, &\qquad D(\eta) = \Omega_\eta, \\ \text{and} &\quad C(\xi, \eta) = 0. \end{aligned} \tag{92}$$

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

Under the Gaussian assumption, the joint density function of ξ and η is

$$N(\xi, \eta) = (2\pi)^{-T} |\Omega_\xi|^{-1/2} |\Omega_\eta|^{-1/2} \exp\big\{ -\frac{1}{2}(\xi'\Omega_\xi^{-1}\xi + \eta'\Omega_\eta^{-1}\eta)\big\}, \tag{93}$$

The problem of estimating ξ and η can be construed as a matter of maximising the likelihood function $N(\xi, \eta)$ subject to the condition that $\xi + \eta = y$. This entails minimising a chi-square criterion function:

$$\begin{aligned} S &= (y - \xi)'\Omega_\eta^{-1}(y - \xi) + \xi'\Omega_\xi^{-1}\xi \\ &= \eta'\Omega_\eta^{-1}\eta + (y - \eta)'\Omega_\xi^{-1}(y - \eta). \end{aligned} \tag{94}$$

The minimising values of ξ and η are, respectively,

$$x = (\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1}\Omega_\eta^{-1}y, \tag{95}$$

$$h = (\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1}\Omega_\xi^{-1}y \tag{96}$$

and it is manifest that $x + h = y$, which is to say that the two estimates obey the same adding-up condition as the true components.

The identities

$$(\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1}\Omega_\eta^{-1} = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}, \tag{97}$$

$$(\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1}\Omega_\xi^{-1} = \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}, \tag{98}$$

which are easily proven by premultiplying and postmultiplying the equations by $\Omega_\xi^{-1} + \Omega_\eta^{-1}$ and $\Omega_\xi + \Omega_\eta$, respectively, can be used to rewrite the estimates as

$$x = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y = Z_\xi y, \tag{99}$$

$$h = \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y = Z_\eta y. \tag{100}$$

The first of these is the formula of (59). It can also be seen, in reference to (92), that

$$x = E(\xi|y) = E(\xi) + C(\xi, y)D^{-1}(y)\{y - E(y)\}, \tag{101}$$

$$h = E(\eta|y) = E(\eta) + C(\eta, y)D^{-1}(y)\{y - E(y)\}, \tag{102}$$

which is to say the estimates are the conditional expectations of the unobserved components—which means that they are also the minimum-mean-square-error estimates.

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

$$\begin{aligned} D(\xi|y) &= D(\xi) - C(\xi, y)D^{-1}(y)C(y, \xi) \\ &= \Omega_\xi - \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\xi, \end{aligned} \tag{103}$$

$$\begin{aligned} D(\eta|y) &= D(\eta) - C(\eta, y)D^{-1}(y)C(y, \eta) \\ &= \Omega_\eta - \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\eta. \end{aligned} \tag{104}$$

These formulae contain the dispersion matrices $D\{E(\xi|y)\} = C(\xi, y)D^{-1}(y)C(y, \xi)$ 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).

Since $y = x + h$, only one of the estimates needs be calculated. The other may be obtained by subtracting the calculated estimate from y. Also, the matrix inversion lemma indicates that

$$\begin{aligned} (\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1} &= \Omega_\eta - \Omega_\eta(\Omega_\eta + \Omega_\xi)^{-1}\Omega_\eta \\ &= \Omega_\xi - \Omega_\xi(\Omega_\eta + \Omega_\xi)^{-1}\Omega_\xi. \end{aligned} \tag{105}$$

Therefore, (103) and (104) represent the same quantity, which is to be expected in view of the adding up.

The identities of (105), which describe the matrix inversion lemma, can be derived from those of (97) and (98). Adding the LHS of (97) to the LHS of (98) gives an identity matrix. Therefore, adding the LHS of (97) to the RHS of (98) also gives the identity matrix:

$$(\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1}\Omega_\eta^{-1} + \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1} = I_T$$

Postmultiplying this by Ω_η and rearranging gives the first of the identities of (105). The other follows by an argument of symmetry.

The filter matrix $Z_\xi = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}$ of (99), which has appeared already under (59), may be regarded as the finite-sample version of the filter formula $\psi_\xi(z) = \gamma^{\xi\xi}(z)/\gamma^{yy}(z)$, of (64). Notice, however, that it is not sufficient merely to replace the argument z within the latter equation by L_T. The reason is that the matrices $\Omega_\xi = \gamma^{\xi\xi}(L_T)$, $\Omega_\eta = \gamma^{\eta\eta}(L_T)$ and $\Omega = \gamma^{yy}(L_T)$ and their inverses fail to commute in multiplication. An order asserts itself amongst the factors of the filter matrices that is immaterial in the case of their infinite-sample analogues.

To investigate the mapping from y to $x = E(\xi|y)$ or, equally, the mapping from y to $h = E(\eta|y)$, we must take account of the various symmetries manifested by the Toeplitz matrices Ω_η and Ω_ξ. The generic Toeplitz matrix Ω is symmetric about the principal (northwest–southeast) diagonal, which is ordinary symmetry. It is symmetric about the secondary (northeast–southwest) diagonal, which is persymmetry. It is invariant with respect to rotations of 180° around the central point at the intersection of its two diagonals, which is centrosymmetry.

Let $H = [e_{T-1}, \ldots, e_1, e_0]$ be the counter-identity matrix, which has units on the secondary diagonal and zeros elsewhere, and let $\Omega^{\#}$ be the counter transpose, which is the reflection of Ω about the secondary diagonal. Then, the symmetries of Ω may be recorded as follows:

$$\begin{array}{clcll} \text{(i)} & \text{Symmetry:} & \Omega & = & \Omega', \\ \text{(ii)} & \text{Persymmetry:} & \Omega & = & \Omega^{\#}, \quad \text{equivalently} \quad H\Omega H = \Omega', \\ \text{(iii)} & \text{Centrosymmetry:} & \Omega & = & (\Omega^{\#})' = \Omega^{\mathrm{r}}, \quad \text{equivalently} \quad H\Omega H = \Omega. \end{array} \tag{106}$$

The matrix of $x = Zy$, which is the estimating equation of the signal ξ, is determined by the equation $\Omega_\xi = Z\Omega$, wherein both Ω_ξ and $\Omega = \Omega_\xi + \Omega_\eta$ are Toeplitz matrices. Therefore, since $H\Omega_\xi H = \Omega_\xi$, $H\Omega H = \Omega$ and $HH = I$, it follows that

$$\Omega_\xi = H\Omega_\xi H = \{HZH\}\{H\Omega H\} = \{HZH\}\Omega = Z\Omega. \tag{107}$$

In view of the nonsingularity of the factors, we conclude from this that $HZH = Z$, which is to say that $Z = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}$ is a centrosymmetric matrix, albeit that it is not a Toeplitz matrix.

Let $y^{\mathrm{r}} = Hy$ and $x^{\mathrm{r}} = Hx$ be y and x in reverse. Then, the centrosymmetric property of Z ensures that both $x = Zy$ and $x^{\mathrm{r}} = Zy^{\mathrm{r}}$. This feature is in accordance with the fact that the direction of time can be reversed without affecting the statistical properties of a stationary process. The use of centrosymmetric matrices in filtering time series has been discussed by Dagum and Luati (2004).

The filter weights that are provided by the rows of the matrices Z vary as the filter progresses through the sample. As the sample size increases, the weights in the central row of Z, when it has an odd number of rows, will tend to the set of constant coefficients that would be derived under the assumption of a doubly-infinite data sequence. These coefficients are symmetric about a central value. The weights of the final row of Z correspond to the coefficients of a one-sided causal filter that looks backwards in time, whereas those in the first row correspond to the same filter looking forwards in time.

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

$$(\Omega_\xi + \Omega_\eta)b = y \tag{108}$$

for the value of b. Thereafter, one can generate

$$x = \Omega_\xi b \qquad \text{and} \qquad h = \Omega_\eta b. \tag{109}$$

If Ω_ξ and Ω_η correspond to the dispersion matrices of moving-average processes, then the solution to equation (108) 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. This is the matrix analogue of a Cramér–Wold factorisation. 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.

8 Filtering Nonstationary Sequences

The problems of filtering a trended data sequence may be overcome by reducing it to stationarity by differencing. The differenced sequence can be filtered and, if necessary, it can be reinflated thereafter to obtain an estimate of a trended data component. If one is seeking to estimate a stationary component of a nonstationary sequence, then the reinflation can be avoided.

It is possible to approach the problem of estimating a trended component by filtering the data directly, without differencing it, provided that sufficient attention is paid to the provision of the necessary initial conditions. This is the preferred approach of some econometricians which leads them to adopt the Kalman filter, which is expounded in section 11. A strong advocacy of the Kalman filter in association with structural time-series models has been made by Durbin and Koopman (2001, §3.5).

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

$$\nabla_T^p = (I - L_T)^p. \tag{110}$$

We may partition the matrix so that $\nabla_T^p = [Q_*, Q]'$, where Q_*' has p rows. The inverse matrix is partitioned conformably to give $\nabla_T^{-p} = [S_*, S]$. We may observe that

$$[\,S_* \quad S\,]\begin{bmatrix} Q_*' \\ Q' \end{bmatrix} = S_*Q_*' + SQ' = I_T, \tag{111}$$

and that

$$\begin{bmatrix} Q_*' \\ Q' \end{bmatrix}[\,S_* \quad S\,] = \begin{bmatrix} Q_*'S_* & Q_*'S \\ Q'S_* & Q'S \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_{T-p} \end{bmatrix}. \tag{112}$$

When the differencing operator is applied to a vector x, the first p elements of the product, which are in d_*, are not true differences and they are liable to be discarded:

$$\nabla_T^p x = \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} y = \begin{bmatrix} d_* \\ d \end{bmatrix}. \tag{113}$$

However, if the elements of d_* are available, then the vector x can be recovered from $d = Q'x$ via the equation

$$x = S_* d_* + Sd. \tag{114}$$

The columns of the matrix S_* provide a basis for the set of polynomials of degree $p - 1$ defined over the integer values $t = 0, 1, \ldots, T - 1$. Therefore, $f = S_* d_*$ is a vector of polynomial ordinates, whilst d_* can be regarded as a vector of p polynomial parameters.

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 sequence of constants, which will be zeros if the degree of the 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

$$\begin{aligned} Q'y &= Q'\xi + Q'\eta \\ &= \delta + \kappa = g, \end{aligned} \tag{115}$$

where

$$\begin{gathered} E(\delta) = \mu\iota, \qquad D(\delta) = \Omega_\delta, \\ E(\kappa) = 0, \qquad D(\kappa) = \Omega_\kappa = Q'\Omega_\eta Q, \\ \text{and} \quad C(\delta, \kappa) = 0. \end{gathered} \tag{116}$$

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

$$\begin{aligned} E(\delta|g) &= E(\delta) + \Omega_\delta(\Omega_\delta + \Omega_\kappa)^{-1}\{g - E(g)\} \\ &= \mu\iota + \Omega_\delta(\Omega_\delta + Q'\Omega_\eta Q)^{-1}\{g - \mu\iota\} = d, \end{aligned} \tag{117}$$

$$\begin{aligned} E(\kappa|g) &= E(\kappa) + \Omega_\kappa(\Omega_\delta + \Omega_\kappa)^{-1}\{g - E(g)\} \\ &= Q'\Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}\{g - \mu\iota\} = k; \end{aligned} \tag{118}$$

and these vectors obey an adding-up condition:

$$Q'y = d + k = g. \tag{119}$$

In (117), 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 (118), 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 drift is p, one can set

$$d = \Omega_\delta(\Omega_\delta + \Omega_\kappa)^{-1}g = \Omega_\delta(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y, \tag{120}$$

$$k = \Omega_\kappa(\Omega_\delta + \Omega_\kappa)^{-1}g = Q'\Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{121}$$

Our object is to recover from d an estimate x of the trend vector ξ via equation (114). The criterion for finding the intial condition or starting value d_* is

$$\text{Minimise} \quad (y - x)'\Omega_\eta^{-1}(y - x) = (y - S_*d_* - Sd)'\Omega_\eta^{-1}(y - S_*d_* - Sd). \tag{122}$$

This requires that the estimated trend x should adhere as closely as possible to the data. The minimising value is

$$d_* = (S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1}(y - Sd). \tag{123}$$

Using this, and defining

$$P_* = S_*(S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1}, \tag{124}$$

we get, from (114), the following value:

$$x = P_*y + (I_T - P_*)Sd. \tag{125}$$

The disadvantage in using this formula directly is that the inverse matrix Ω_η^{-1}, which is of order T, is liable to have nonzero elements in every location. The appropriate recourse is to use the identity

$$\begin{aligned} I_T - P_* &= I_T - S_*(S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1} \\ &= \Omega_\eta Q(Q'\Omega_\eta Q)^{-1}Q' \end{aligned} \tag{126}$$

to provide an alternative expression for the projection matrix $I_T - P_*$ that incorporates the narrow-band matrix Ω_η instead of its inverse. The equality follows from the fact that, if $\text{Rank}[R, S_*] = T$ and if $S_*'\Omega_\eta^{-1}R = 0$, then

$$I_T - S_*(S_*'\Omega_\eta^{-1}S_*)^{-1}S_*'\Omega_\eta^{-1} = R(R'\Omega_\eta^{-1}R)^{-1}R'\Omega_\eta^{-1}. \tag{127}$$

Setting $R = \Omega_\eta Q$ gives the result. Given that $x = y - h$, it follows that we can write

$$\begin{aligned} x &= y - (I_T - P_*)Sk \\ &= y - \Omega_\eta Q(Q'\Omega_\eta Q)^{-1}k, \end{aligned} \tag{128}$$

where the second equality depends upon $Q'S = I_T$. On substituting k from (121) into the equation of (128), we get

$$x = y - \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{129}$$

Equation (129) can also be derived via a straightforward generalisation of the chi-square criterion of (94). 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 (94) and replacing the inverse of Ω_ξ^{-1} by the generalised inverse $\Omega_\xi^+ = Q\Omega_\delta^{-1}Q'$, we get the function

$$S = (y - \xi)'\Omega_\eta^{-1}(y - \xi) + \xi' Q\Omega_\delta^{-1}Q'\xi, \tag{130}$$

of which the minimising value is

$$x = (Q\Omega_\delta^{-1}Q' + \Omega_\eta^{-1})^{-1}\Omega_\eta^{-1}y. \tag{131}$$

The matrix inversion lemma gives

$$(Q\Omega_\delta^{-1}Q' + \Omega_\eta^{-1})^{-1} = \Omega_\eta - \Omega_\eta Q(Q'\Omega_\eta Q + \Omega_\delta)^{-1}Q'\Omega_\eta; \tag{132}$$

and putting this into (131) gives the expression under (129). The matrix of (132) also constitutes the error dispersion matrix $D(\eta|y) = D(\xi|y)$ which, in view of their adding-up property, is common to the estimates of the two components.

At this point, we may observe that it is possible to estimate two independent nonstationary components ξ and η from their combined data sequence $y = \xi + \eta$. Define matrices ∇_ξ and ∇_η such that

$$\nabla_\xi \xi = \begin{bmatrix} Q_{\xi*}' \\ Q_\xi' \end{bmatrix}\xi = \begin{bmatrix} \delta_* \\ \delta \end{bmatrix} \quad \text{and} \quad \nabla_\eta \eta = \begin{bmatrix} Q_{\eta*}' \\ Q_\eta' \end{bmatrix}\eta = \begin{bmatrix} \kappa_* \\ \kappa \end{bmatrix}, \tag{133}$$

and assume that $\text{Rank}[Q_\xi, Q_\eta] = T$. The operators Q_ξ' and Q_ξ' reduce the respective components to independent stationary mean-zero sequences δ and κ, with $E(\delta) = 0$, $D(\delta) = \Omega_\delta$ and $E(\kappa) = 0$, $D(\kappa) = \Omega_\kappa$.

Then, an appropriate criterion for finding the estimates of the original components is to minimise the function

$$\begin{aligned} S &= (y - \xi)'Q_\eta\Omega_\kappa^{-1}Q_\eta'(y - \xi) + \xi' Q_\xi\Omega_\delta^{-1}Q_\xi'\xi \\ &= \eta' Q_\eta\Omega_\kappa^{-1}Q_\eta'\eta + (y - \eta)'Q_\xi\Omega_\delta^{-1}Q_\xi'(y - \eta) \end{aligned} \tag{134}$$

in respect of ξ and η. This gives rise to the following equations:

$$\begin{aligned} x &= (Q_\xi\Omega_\delta^{-1}Q'_\xi + Q_\eta\Omega_\kappa^{-1}Q'_\eta)^{-1}Q_\eta\Omega_\kappa^{-1}Q'_\eta y, \\ h &= (Q_\xi\Omega_\delta^{-1}Q'_\xi + Q_\eta\Omega_\kappa^{-1}Q'_\eta)^{-1}Q_\xi\Omega_\delta^{-1}Q'_\xi y. \end{aligned} \tag{135}$$

Since neither $Q_\xi\Omega_\delta^{-1}Q'_\xi$ nor $Q_\eta\Omega_\kappa^{-1}Q'_\eta$ is invertible, the matrix inversion lemma is no longer applicable and, therefore, computationally efficient forms that exploit the ease of inverting narrow-band Toeplitz matrices are not directly available. Nevertheless, McElroy (2006) has demonstrated a practical implementation of the above formulae.

It is uncommon of find a model that lacks a stationary noise component. However, it is quite common to find a nonstationary component that comprises two nonstationary subcomponents that require to be separated. In that case, the noise component may be lumped together with one of the nonstationary components to enable the estimating equations of (135) to be exploited. Then, the noise can be separated from the component with which it has been combined. Alternatively, the composite nonstationary component can be separated from the stationary noise, wherafter it can be decomposed into its constituent components in a manner that requires starting values to be estimated explicitly.

Example. A typical model of an econometric time series, described by the equation

$$\begin{aligned} y &= \xi + \eta \\ &= (\tau + \sigma) + \eta, \end{aligned} \tag{136}$$

comprises a trend/cycle component τ and a seasonal component σ that are described by ARIMA models with real and complex unit roots respectively. The remaining component η is irregular white noise. The models of these components may have been obtained, for example, by applying the principle of canonical decompositions, which is to be described in section 10, to an aggregate model of the data sequence.

To reduce the data to stationarity, an operator is used that is the product of a detrending operator $\nabla_\tau = (I - L_T)^p$ and a deseasonalising operator $\nabla_\sigma = (I - L_T^q)(I - L_T)^{-1} = I + L_T + \cdots + L_T^{q-1}$, where q is the number of seasons (or months). (The matrix ∇_σ, which corresponds to a seasonal summation operator, is used instead of the seasonal differencing operator $I - L_T^q$ because it can be assumed, without loss of generality, that the seasonal deviations from the trend have zero mean.)

Let the product of the two operators be denoted by $\nabla_\tau\nabla_\sigma = \nabla_\xi = [Q_*, Q]'$, where Q'_* contains the first $p+q-1$ rows of the matrix, and let the inverse operator $\nabla_\xi^{-1} = \Sigma = [S_*, S]$ be partitioned conformably such that S_* contains the first $p+q-1$ columns. The factors $\Sigma_\tau = \nabla_\tau^{-1}$ and $\Sigma_\sigma = \nabla_\sigma^{-1}$ of Σ are further partitioned as $\Sigma_\tau = [S_{\tau *}, S_\tau]$ and $\Sigma_\sigma = [S_{\sigma *}, S_\sigma]$.

Let the components of the transformed data be denoted by $Q'\xi = \delta$, $Q'\tau = \delta_\tau$ and $Q'\sigma = \delta_\sigma$. Then, there is

$$\begin{aligned} Q'y &= Q'\xi + Q'\eta \\ &= Q'(\tau + \sigma) + \kappa = (\delta_\tau + \delta_\sigma) + \kappa. \end{aligned} \tag{137}$$

Also, let the estimates of τ and σ be denoted by r and s and those of δ_τ and δ_σ by d_τ and d_r. Then, in parallel with equation (137), there is

$$\begin{aligned} Q'y &= Q'x + Q'h \\ &= Q'(r + s) + k = (d_\tau + d_\sigma) + k. \end{aligned} \tag{138}$$

The estimates d_τ, d_σ and k may be obtained from the transformed data $g = Q'y$ by a process of linear filtering. Then, it is required to form r, s and h from these elements.

First, consider

$$\begin{aligned} x = (r + s) &= S_*d_* + Sd \\ &= S_*d_* + S(d_\tau + d_\sigma). \end{aligned} \tag{139}$$

Here, d_* is computed according the formula of (123). Given x, an estimate $h = y - x$ of the irregular component can be formed. Next, there is an equation

$$S_* d_* = \begin{bmatrix} S_{\tau *} & S_{\sigma *} \end{bmatrix} \begin{bmatrix} d_{\tau *} \\ d_{\sigma *} \end{bmatrix}. \tag{140}$$

This may be solved uniquely for $d_{\tau *}$ and $d_{\sigma *}$; and, for this purpose, only the first $p + q - 1$ rows of the system are required. Thereafter, the estimates of τ and σ are given by

$$r = S_{\tau *} d_{\tau *} + S d_\tau \quad \text{and} \quad s = S_{\sigma *} d_{\sigma *} + S d_\sigma . \tag{141}$$

What has been recounted in this example is, essentially, the method proposed by Bell (1984).

9 Filtering in the Frequency Domain

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

$$\begin{aligned} \Omega_\xi^\circ = \bar{U}\gamma_\xi(D)U, \quad \Omega_\eta^\circ = \bar{U}\gamma_\eta(D)U \quad &\text{and} \\ \Omega^\circ = \Omega_\xi^\circ + \Omega_\eta^\circ = \bar{U}\{\gamma_\xi(D) + \gamma_\eta(D)\}U&, \end{aligned} \tag{142}$$

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 (99) and (100) by their circulant counterparts, we derive the following formulae:

$$x = \bar{U}\gamma_\xi(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}Uy = P_\xi y, \tag{143}$$

$$h = \bar{U}\gamma_\eta(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}Uy = P_\eta y. \tag{144}$$

Similar replacements within the formulae (103) and (104) 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 $T = 2^q$, where q is some integer.)

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 (143) gives

$$x = \bar{U}\Lambda_\xi\{\Lambda_\xi + \Lambda_\eta\}^+ Uy = \bar{U}P_\xi Uy, \tag{145}$$

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

$$\Omega_\xi^\circ - \Omega_\xi^\circ(\Omega_\xi^\circ + \Omega_\eta^\circ)^+\Omega_\xi^\circ = \bar{U}\Lambda_\xi U - \bar{U}\Lambda_\xi(\Lambda_\xi + \Lambda_\eta)^+\Lambda_\xi U. \tag{146}$$

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 $x = \xi$, and the signal is recovered perfectly.

In applying the Fourier method of signal extraction to non-stationary sequences, it is necessary to reduce the data to stationarity. The reduction can be achieved by the differencing operation represented by equation (115). The components δ and κ of the differenced data may be estimated via the equations

$$\begin{aligned} d &= \bar{U}\Lambda_\delta(\Lambda_\delta+\Lambda_\kappa)^+Ug = P_\delta g, && (147)\\ k &= \bar{U}\Lambda_\kappa(\Lambda_\delta+\Lambda_\kappa)^+Ug = P_\kappa g. && (148)\end{aligned}$$

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

Whereas these 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 by cumulating d and k in the manner of equation (114). The procedure, which originates in the time-domain approach, requires explicit initial conditions, denoted by d_* and k_*.

It may also be appropriate, in this context, to replace the criteria of (122), which generates the values of d_*, by simplified criterion wherein Ω_η is replaced by the identity matrix I_T. A similar criterion can be used for finding a value for k_* within the equation $h = S_*k_* + Sk$. Then,

$$d_* = (S_*'S_*)^{-1}S_*'(y - Sd), \quad \text{and} \quad k_* = (S_*'S_*)^{-1}S_*'Sk. \tag{149}$$

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 of 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 highpass or bandpass component is given by the equation

$$h = \bar{U}J_\psi Ug. \tag{150}$$

10 Structural Time Series Models

In economics, it is traditional to decompose time series into a variety of components, some or all of which may be present in a particular instance.

One is liable to assume that the relative proportions of the components of an aggregate index are maintained, approximately, in spite of the variations in their levels. Therefore, the basic model of an economic index is a multiplicative one; and, if $Y(t)$ is the sequence of values of an economic index, then it can be expressed as

$$Y(t) = L(t) \times C(t) \times S(t) \times H(t), \tag{151}$$

where

$L(t)$ is the global trend,
$C(t)$ is a secular cycle,
$S(t)$ is the seasonal variation and
$H(t)$ is an irregular component.

Many of the more prominent macroeconomic indicators are amenable to a decomposition of this sort. One can imagine, for example, a quarterly index of Gross Domestic Product which appears to be following an exponential growth trend $L(t)$.

The trend might be obscured, to some extent, by a superimposed cycle $C(t)$ with a period of roughly four and a half years, which happens to correspond, more or less, to the average lifetime of the legislative assembly. The reasons for this curious coincidence need not concern us here.

The ghost of an annual cycle $S(t)$ might also be apparent in the index; and this could be a reflection of the fact that some economic activities, such as building construction, are affected significantly by the weather and by the duration of sunlight.

When the foregoing components—the trend, the secular cycle and the seasonal cycle—have been extracted from the index, the residue should correspond to an irregular component $H(t)$ for which no unique explanation can be offered.

The logarithms $y(t) = \ln Y(t)$ of the aggregate index are amenable to an additive decomposition. Thus, equation (151) gives rise to

$$\begin{aligned} y(t) &= \{\lambda(t) + \gamma(t)\} + \sigma(t) + \eta(t) \\ &= \tau(t) + \sigma(t) + \eta(t), \end{aligned} \tag{152}$$

where $\lambda(t) = \ln L(y)$, $\gamma(t) = \ln C(t)$, $\sigma(t) = \ln S(t)$ and $\eta(t) = \ln H(t)$. Since the trend and the cycles are not easily separable, there is a case for combining them in a component $T(t) = L(t)\times C(t)$, of which the logarithm is $\ln T(t) = \tau(t)$.

In the structural time-series model, the additive components are modelled by independent ARMA or ARIMA process. Thus

$$\begin{aligned} y(z) &= \tau(z) + \sigma(z) + \eta(z) \\ &= \frac{\theta_\tau(z)}{\phi_\tau(z)}\zeta_\tau(z) + \frac{\theta_\sigma(z)}{\phi_\sigma(z)}\zeta_\sigma(z) + \eta(z), \end{aligned} \tag{153}$$

where $\zeta_\tau(z)$, $\zeta_\sigma(z)$ and $\eta(z)$ are the z-transforms of statistically independent white-noise processes. Within the autoregressive polynomial $\phi_\tau(z)$ of the trend component will be found the unit-root factor $(1-z)^p$, whereas the autoregressive polynomial $\phi_\sigma(z)$ of the seasonal component will contain the factor $(1 + z + \cdots + z^{s-1})^D$, wherein s stands for the number of periods in a seasonal cycle.

The sum of a set of ARIMA processes is itself and ARIMA process. Therefore, $y(t)$ can be expressed as a univariate ARIMA process which is described as the reduced form of the time-series model:

$$y(z) = \frac{\theta(z)}{\phi(z)}\varepsilon(z) = \frac{\theta(z)}{\phi_\sigma(z)\phi_\tau(z)}\varepsilon(z). \tag{154}$$

Here, $\varepsilon(z)$ stands for the z-transform of a synthetic white-noise process.

There are two alternative approaches to the business of estimating the structural model and of extracting its components. The first approach, which is described as the canonical approach, is to estimate the parameters of the reduced-form ARIMA model. From these parameters, the Wiener–Kolmogorov filters that are appropriate for extracting the components can be constructed.

On the assumption that the degree of the moving-average polynomial $\theta(z)$ is at least equal to that of the autoregressive polynomial $\phi(z)$, there is a partial-fraction decomposition of the autocovariance generating function of the model into three components, which correspond to the trend effect, the seasonal effect and an irregular influence. Thus

$$\frac{\theta(z)\theta(z^{-1})}{\phi_\sigma(z)\phi_\tau(z)\phi_\tau(z^{-1})\phi_\sigma(z^{-1})} = \frac{Q_\tau(z)}{\phi_\tau(z)\phi_\tau(z^{-1})} + \frac{Q_\sigma(z)}{\phi_\sigma(z)\phi_\sigma(z^{-1})} + R(z). \tag{155}$$

Here, the first two components on the RHS represent proper rational fractions, whereas the irregular component $R(z)$ is an ordinary polynomial. If the degree of the moving-average polynomial in the reduced form is less than that of the autoregressive polynomial, then the irregular component is missing from the decomposition in the first instance.

To obtain the spectral density function $f(\omega)$ of $y(t)$ and of its components, we set $z = e^{-\mathrm{i}\omega}$ in (155). (This function is more properly described as a pseudo-spectrum in view of the singularities occasioned by the unit roots in the denominators of the first two components.) The spectral decomposition can be written as

$$f(\omega) = f_\tau(\omega) + f_\sigma(\omega) + f_R(\omega). \tag{156}$$

Let $\nu_\tau = \min\{f_\tau(\omega)\}$ and $\nu_\sigma = \min\{f_\sigma(\omega)\}$. These are the elements of white noise embedded in $f_\tau(\omega)$ and $f_\sigma(\omega)$. The principle of canonical decomposition is that the white-noise elements should be reassigned to the residual component. (The principle of canonical decompositions has been expounded, for example, by Hillmer and Tiao (1982), Maravall and Pierce (1987), and, more recently, Kaiser and Maravall (2001).) On defining

$$\begin{aligned} \gamma_\tau(z)\gamma_\tau(z^{-1}) &= Q_\tau(z) - \nu_\tau\phi_\tau(z)\phi_\tau(z^{-1}), \\ \gamma_\sigma(z)\gamma_\sigma(z^{-1}) &= Q_\sigma(z) - \nu_\sigma\phi_\sigma(z)\phi_\sigma(z^{-1}), \\ \text{and}\quad \rho(z)\rho(z^{-1}) &= R(z) + \nu_\tau + \nu_\sigma, \end{aligned} \tag{157}$$

the canonical decomposition of the generating function can be represented by

$$\frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})} = \frac{\gamma_\tau(z)\gamma_\tau(z^{-1})}{\phi_\tau(z)\phi_\tau(z^{-1})} + \frac{\gamma_\sigma(z)\gamma_\sigma(z^{-1})}{\phi_\sigma(z)\phi_\sigma(z^{-1})} + \rho(z)\rho(z^{-1}). \tag{158}$$

There are now two improper rational functions on the RHS, which have equal degrees in their numerators and denominators.

According to Wiener–Kolmogorov theory, the optimal signal-extraction filter for the trend component is

$$\begin{aligned} \beta_\tau(z) &= \frac{\gamma_\tau(z)\gamma_\tau(z^{-1})}{\phi_\tau(z)\phi_\tau(z^{-1})} \times \frac{\phi_\sigma(z)\phi_\tau(z)\phi_\tau(z^{-1})\phi_\sigma(z^{-1})}{\theta(z)\theta(z^{-1})} \\ &= \frac{\gamma_\tau(z)\gamma_\tau(z^{-1})\phi_\sigma(z)\phi_\sigma(z^{-1})}{\theta(z)\theta(z^{-1})}. \end{aligned} \tag{159}$$

This has the form of the ratio of the autocovariance generating function of the trend component to the autocovariance generating function of the process $y(t)$.

Observe that, in the process of forming this filter, the factor $\phi_\tau(z)\phi_\tau(z^{-1})$ is cancelled out. With the consequent removal of the unit-root factor $(1-z)^p(1-z^{-1})^p$ from the denominator, the basis of a stable filter is created which, with the provision of appropriate starting values, can be applied to nonstationary data. This filter would also serve to extract a differenced version of the component $\tau(t)$ from the differenced data. The filter that serves to extract the seasonal component is of a similar construction.

These formulations presuppose a doubly-infinite data sequence; and they must be translated into a form that can be implemented with finite sequences. The various ways of achieving this have been described in section 5; and, in the TRAMO–SEATS program of Gómez and Maravall (1996) and of Caporello and Maravall (2004), the contragrade method of Burman (1980) has been adopted, which entails a unique treatment of the start-up problem.

The alternative method of estimating the parameters of the structural model and of extracting the unobserved components makes use of the fact that a univariate autoregressive moving-average model can be expressed as a first-order multivariate Markov model, which constitutes a state-space representation of the model. This allows the structural parameters to be estimated directly, as opposed to being inferred indirectly from the parameters of the reduced-form model.

The state-space approach to the structural time-series model was pioneered by Harrison and Stevens (1971, 1976). An extensive account of the approach has been provided by Harvey (1989). Other important references are the books of West and Harrison (1997) and Kitagawa and Gersch (1996). Proietti (2002) has also provided a brief but thorough account. A brief introductory survey has been provided by West (1997), and an interesting biomedical application has been demonstrated by West *et al.* (1999).

The methods may be illustrated by considering the so-called basic structural model, which has been popularised by Harvey (1989). The model, which lacks a non-seasonal cyclical component, can be subsumed under the second of the equations of (152).

The trend or levels component $\tau(t)$ of this model is described by a stochastic process that generates a trajectory that is approximately linear within a limited locality. Thus

$$\begin{gathered} \tau(t) = \tau(t-1) + \beta(t-1) + \upsilon(t) \quad \text{or, equivalently,} \\ \nabla(z)\tau(z) = z\beta(z) + \upsilon(z), \end{gathered} \tag{160}$$

where $\nabla(z) = 1 - z$ is the difference operator. That is to say, the change in the level of the trend is compounded from the slope parameter $\beta(t-1)$, generated in the previous period, and a small white-noise disturbance $\upsilon(t)$. The slope parameter follows a random walk. Thus

$$\beta(t) = \beta(t-1) + \zeta(t) \quad \text{or, equivalently,} \quad \nabla(z)\beta(z) = \zeta(z), \tag{161}$$

where $\zeta(t)$ denotes a white-noise process that is independent of the disturbance process $\upsilon(t)$. By applying the difference operator to equation (160) and substituting from (161), we find that

$$\begin{aligned} \nabla^2(z)\tau(z) &= \nabla(z)z\beta(z) + \nabla(z)\upsilon(z) \\ &= z\zeta(z) + \nabla(z)\upsilon(z). \end{aligned} \tag{162}$$

The two terms of the RHS can be combined to form a first-order moving-average process, whereupon the process generating $\tau(t)$ can be described by an integrated moving-average IMA(2, 1) model. Thus

$$\begin{aligned} \nabla^2(z)\tau(z) &= z\zeta(z) + \nabla(z)\upsilon(z) \\ &= (1 - \mu z)\varepsilon(z). \end{aligned} \tag{163}$$

A limiting case arises when the variance of the white-noise process $\zeta(t)$ in equation (161) tends to zero. Then, the slope parameter tends to a constant β, and the process by which the trend is generated, which has been identified as an IMA(2,1) process, becomes a random walk with drift.

Another limiting case arises when the variance of $\upsilon(t)$ in equation (160) tends to zero. Then, the overall process generating the trend becomes a second-order random walk, and the resulting trends are liable to be described as smooth trends. When the variances of $\zeta(t)$ and $\upsilon(t)$ are both zero, then the process $\tau(t)$ degenerates to a simple linear time trend.

The seasonal component of the structural time-series model is described by the equation

$$\begin{aligned} &\sigma(t) + \sigma(t-1) + \cdots + \sigma(t-s+1) = \omega(t) \\ &\text{or, equivalently,} \\ &S(z)\sigma(z) = \omega(z), \end{aligned} \tag{164}$$

where $S(z) = 1 + z + z^2 + \cdots + z^{s-1}$ is the seasonal summation operator, s is the number of observation per annum and $\omega(t)$ is a white-noise process.

The equation implies that the sum of s consecutive values of this component will be a random variable distributed about a mean of zero. To understand this construction, we should note that, if the seasonal pattern were perfectly regular and invariant, then the sum of the consecutive values would be identically zero. Since the sum is a random variable with a zero mean, some variability can occur in the seasonal pattern.

By substituting equations (162) and (164) into equation (152), we seen that the structural model can be represented by the equation

$$\begin{aligned} &\nabla^2(z)S(z)y(z) = S(z)z\zeta(z) + \nabla(z)S(z)\upsilon(z) + \nabla^2(z)\omega(z) + \nabla^2(z)S\eta(z), \\ &\text{or, equivalently,} \\ &\nabla(z)\nabla_s(z)y(z) = S(z)z\zeta(z) + \nabla_s(z)\upsilon(z) + \nabla^2(z)\omega(z) + \nabla(z)\nabla_s(z)\eta(z), \end{aligned} \tag{165}$$

where $\zeta(t)$, $\upsilon(t)$, $\omega(t)$ and $\eta(t)$ are mutually independent white-noise processes. Here, the alternative expression comes from using the identity

$$\nabla(z)S(z) = (1-z)(1 + z + \cdots + z^{s-1}) = (1 - z^s) = \nabla_s(z).$$

We should observe that the RHS or equation (165) corresponds to a moving average of degree $s + 1$, which is typically subject to a number of restriction on its parameters. The restrictions arise from the fact there are only four parameters in the model of (165), which are the white-noise variances $V\{\zeta(t)\}$, $V\{\upsilon(t)\}$, $V\{\omega(t)\}$ and $V\{\eta(t)\}$, whereas there are $s + 1$ moving-average

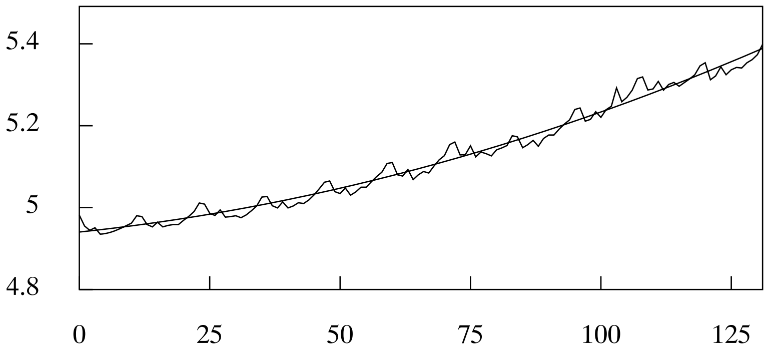

Figure 9: The plot of 132 monthly observations on the U.S. money supply, beginning in January 1960. A quadratic function has been interpolated through the data.

parameters and a variance parameter in the unrestricted reduced-form of the seasonal ARIMA model.

The basic structural model can be represented is a state-space form which comprises a transition equation, which constitutes a first-order vector autoregressive process, and an accompanying measurement equation. For notational convenience, let $s = 4$, which corresponds to the case of quarterly observations. Then, the transition equation, which gathers together equations (160), (161) and (164), is

$$\begin{bmatrix} \tau(t) \\ \beta(t) \\ \sigma(t) \\ \sigma(t-1) \\ \sigma(t-2) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tau(t-1) \\ \beta(t-1) \\ \sigma(t-1) \\ \sigma(t-2) \\ \sigma(t-3) \end{bmatrix} + \begin{bmatrix} \upsilon(t) \\ \zeta(t) \\ \omega(t) \\ 0 \\ 0 \end{bmatrix}. \tag{166}$$

The observation equation, which corresponds to (152), is

$$y(t) = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tau(t) \\ \beta(t) \\ \sigma(t) \\ \sigma(t-1) \\ \sigma(t-2) \end{bmatrix} + \eta(t). \tag{167}$$

The state-space model is amenable to the Kalman filter and the associated smoothing algorithms, which can be used in estimating the parameters of the model and in extracting estimates of the unobserved components $\tau(t)$, $\sigma(t)$.

There are various ways of handling, within the context of the Kalman filter, the start-up problem that is associated with filtering of nonstationary data sequences. These will be touched upon at the end of the next section.

Example. Figure 9 shows the logarithms of a monthly sequence of 132 observations of the U.S. money supply, through which a quadratic function has been interpolated. This provides a simple way of characterising the growth over the period in question.

However, it is doubtful whether such an analytic function can provide an adequate representation of a trend that is subject to irregular variations; and we prefer to estimate the trend more flexibly by applying a linear filter to the data. In order to devise an effective filter, it is helpful to know the extent of the frequency band in which the spectral effects of the trend are located.

It is difficult to discern the spectral structure of the data in the periodogram of the trended sequence y. This is dominated by the effects of the disjunctions in the periodic extension of the data that occur where the end of one replication of the data sequence joins the beginning of the next. In fact, the periodic extension of a segment of a linear trend will generate a sawtooth function,

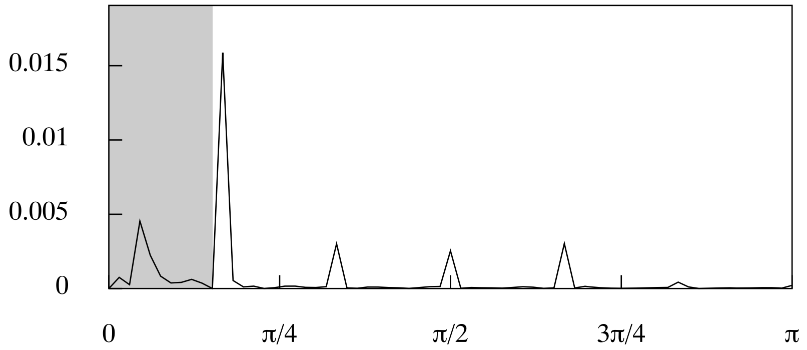

Figure 10: The periodogram of the residuals of the logarithmic money-supply data.

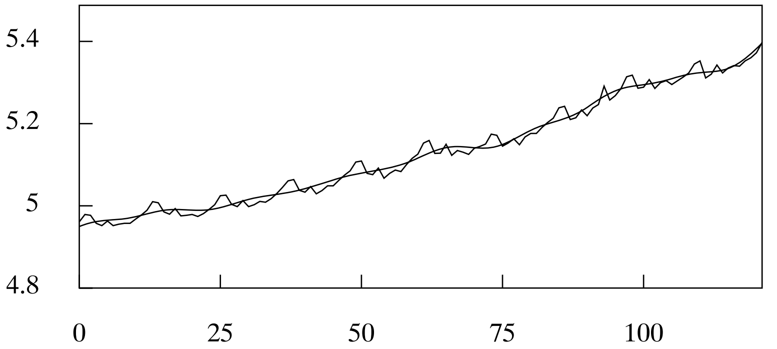

Figure 11: logarithms of 132 monthly observations on the U.S. money supply, beginning in January 1960. A trend, estimated by the Fourier method, has been interpolated through the data.

of which the periodogram will have the form of a rectangular hyperbola, within which any finer spectral detail will be concealed.

On the other hand, if a d-fold differencing operation is used to reduce the data to stationarity to produce $g = Qy$, then one may find that the low-frequency spectral ordinates have been diminished to such an extent that the structure of the trend has become invisible. The problem will to be exacerbated when the data contain a strong seasonal component, which may be amplified by the differencing operation to become the dominant feature of the periodogram.

An effective way of discerning the spectral structure of the data is to examine the periodograms of the residuals obtained by fitting polynomials of various degrees to the data. The residual sequence from fitting a polynomial of degree d, can expressed as

$$r = Q(Q'Q)^{-1}Q'y, \tag{168}$$

where Q' is the aforementioned differencing operator. This sequence contains the same information as the differenced sequence $g = Q'y$, but its periodogram renders the spectral structure visible over the entire frequency range.

Figure 10 which shows the periodogram of the residuals from the quadratic detrending of Figure 9. There is a significant spectral mass within the frequency range $[0, \pi/6)$, of which the upper bound is the fundamental frequency of the seasonal fluctuations. This mass properly belongs to the trend and, if the trend had been adequately estimated, it would not be present in the periodogram of the residuals.

To construct a better estimate of the trend, an ideal lowpass filter, with a sharp cut-off frequency a little short of $\pi/6$, has been applied to the twice differenced data and the filtered sequence has been reinflated with initial conditions that are supplied by equation (123). The result is the trend that is shown in Figure 11. The pass band of the ideal lowpass filter has been superimposed upon

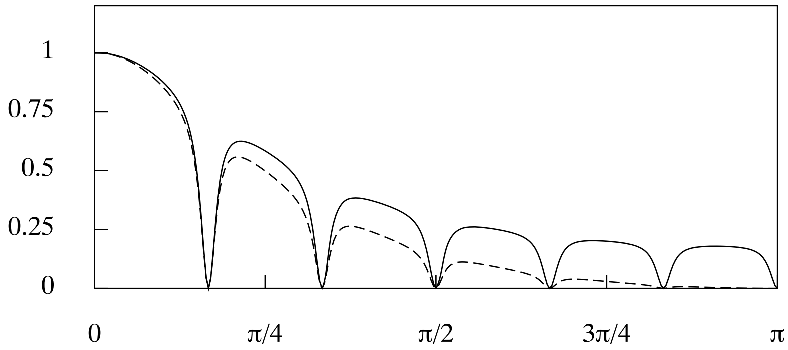

Figure 12: The gain function of the trend-extraction filter obtained from the STAMP program (solid line) together with that of the canonical trend-extraction filter (broken line) obtained from the TRAMO–SEATS program.

the periodogram of Figure 10 as a shaded area.

Figure 12 shows the gains of the trend estimation filters that have been obtained by applying two of the model-based procedures to the data. The outer envelope is the gain of a trend extraction filter obtained in the process of using the STAMP program to estimate the components of the data. The inner envelope represents the gain of the analogous filter from the TRAMO–SEATS program. The indentations in the gain functions of both filters at the frequencies $\pi j/6; j = 1, \ldots, 6$ have the effect of nullifying the seasonal elements and of preventing them from entering the trend.

The two model-based filters differ greatly from the ideal filter. Disregarding the indentations, one can see how the gain of the filters is reduced only gradually as the frequency value increases. The trend component extracted by the STAMP filter would contain a substantial proportion of the non-seasonal high-frequency components that are present in the original data.

In practice, however, the trends that are estimated by the ideal filter and by the two model-based filters are virtually indistinguishable in the case of the money supply data. The reason for this is that, after the elimination of the seasonal components, whether it be by nullifying all elements of frequencies in excess of $\pi/6$ or only by eliminating the elements in the vicinities of the seasonal frequencies of $\pi j/6; j = 1, \ldots, 6$, there is virtually nothing remaining in the data but the trend. Therefore, in this case, the potential of the two model-based filters to transmit high-frequency components can do no harm.

In other cases, it has been observed that the STAMP filter produces a trend estimate that has a profile which is noticeably rougher than the one produced by the TRAMO–SEATS program—see Pollock (2002), for example—and this is a testimony to fact that the latter program, which observes the so-called canonical principle, suppresses the high-frequency noise more emphatically.

11 The Kalman Filter and the Smoothing Algorithm

One of the reasons for setting a structural time-series model in a state-space form is to make it amenable to the application the Kalman filter, which may be used both for estimating the parameters of the model and for extracting the unobserved components. To obtain estimates that take full advantage of all of the sampled data, a smoothing algorithm must also be deployed. These algorithms are described in the present section.

The state-space model, which underlies the Kalman filter, consists of two equations

$$y_t = H\xi_t + \eta_t, \qquad \textit{Observation Equation} \tag{169}$$

$$\xi_t = \Phi\xi_{t-1} + \nu_t, \qquad \textit{Transition Equation} \tag{170}$$

where y_t is the observation on the system and ξ_t is the state vector. The observation error η_t and the state disturbance ν_t are mutually uncorrelated random vectors of zero mean with dispersion matrices

$$D(\eta_t) = \Omega \qquad \text{and} \qquad D(\nu_t) = \Psi. \tag{171}$$

It is assumed that the matrices H, Φ, Ω and Ψ are known and that an initial estimate x_0 is available for the state vector ξ_0 at time $t = 0$ together with a dispersion matrix $D(\xi_0) = P_0$. This set of initial information is denoted by $\mathcal{I}_0$. (In a more general formulation, the parameter matrices would be allowed to vary with time, but here they are constant.) The information available at time t is $\mathcal{I}_t = \{y_t, \ldots, y_1, \mathcal{I}_0\} = \{y_t, \mathcal{I}_{t-1}\}$.

The Kalman-filter equations determine the state-vector estimates $x_{t|t-1} = E(\xi_t|\mathcal{I}_{t-1})$ and $x_t = E(\xi_t|\mathcal{I}_t)$ and their associated dispersion matrices $P_{t|t-1}$ and P_t from the values x_{t-1}, P_{t-1} of the previous period. From $x_{t|t-1}$, the prediction $\hat{y}_{t|t-1} = Hx_{t|t-1}$ is formed which has a dispersion matrix F_t. A summary of these equations is as follows:

$$\begin{aligned}
x_{t|t-1} &= \Phi x_{t-1}, & &\textit{State Prediction} & (172)\\
P_{t|t-1} &= \Phi P_{t-1}\Phi' + \Psi, & &\textit{Prediction Dispersion} & (173)\\
e_t &= y_t - Hx_{t|t-1}, & &\textit{Prediction Error} & (174)\\
F_t &= HP_{t|t-1}H' + \Omega, & &\textit{Error Dispersion} & (175)\\
K_t &= P_{t|t-1}H'F_t^{-1}, & &\textit{Kalman Gain} & (176)\\
x_t &= x_{t|t-1} + K_t e_t, & &\textit{State Estimate} & (177)\\
P_t &= (I - K_t H)P_{t|t-1}. & &\textit{Estimate Dispersion} & (178)
\end{aligned}$$

The equations of the Kalman filter may be derived using the ordinary algebra of conditional expectations which indicates that, if x, y are jointly distributed variables which bear the linear relationship $E(y|x) = \alpha + B\{x - E(x)\}$, then

$$E(y|x) = E(y) + C(y,x)D^{-1}(x)\big\{x - E(x)\big\}, \tag{179}$$

$$D(y|x) = D(y) - C(y,x)D^{-1}(x)C(x,y), \tag{180}$$

$$E\big\{E(y|x)\big\} = E(y), \tag{181}$$

$$D\big\{E(y|x)\big\} = C(y,x)D^{-1}(x)C(x,y), \tag{182}$$

$$D(y) = D(y|x) + D\big\{E(y|x)\big\}, \tag{183}$$

$$C\big\{y - E(y|x), x\big\} = 0. \tag{184}$$

Of the equations listed under (172)—(178), those under (174) and (176) are merely definitions. To demonstrate equation (172), we use (181) to show that

$$\begin{aligned}
E(\xi_t|\mathcal{I}_{t-1}) &= E\big\{E(\xi_t|\xi_{t-1})|\mathcal{I}_{t-1}\big\}\\
&= E\big\{\Phi\xi_{t-1}|\mathcal{I}_{t-1}\big\}\\
&= \Phi x_{t-1}.
\end{aligned} \tag{185}$$

We use (183) to demonstrate equation (173):

$$\begin{aligned}
D(\xi_t|\mathcal{I}_{t-1}) &= D(\xi_t|\xi_{t-1}) + D\big\{E(\xi_t|\xi_{t-1})|\mathcal{I}_{t-1}\big\}\\
&= \Psi + D\big\{\Phi\xi_{t-1}|\mathcal{I}_{t-1}\big\}\\
&= \Psi + \Phi P_{t-1}\Phi'.
\end{aligned} \tag{186}$$

To obtain equation (175), we substitute (169) into (174) to give $e_t = H(\xi_t - x_{t|t-1}) + \eta_t$. Then, in view of the statistical independence of the terms on the RHS, we have

$$\begin{aligned}
D(e_t) &= D\big\{H(\xi_t - x_{t|t-1})\big\} + D(\eta_t)\\
&= HP_{t|t-1}H' + \Omega = D(y_t|\mathcal{I}_{t-1}).
\end{aligned} \tag{187}$$

To demonstrate the updating equation (177), we begin by noting that

$$\begin{aligned}
C(\xi_t, y_t|\mathcal{I}_{t-1}) &= E\big\{(\xi_t - x_{t|t-1})y_t'\big\}\\
&= E\big\{(\xi_t - x_{t|t-1})(H\xi_t + \eta_t)'\big\}\\
&= P_{t|t-1}H'.
\end{aligned} \tag{188}$$

It follows from (179) that

$$\begin{aligned} E(\xi_t|\mathcal{I}_t) &= E(\xi_t|\mathcal{I}_{t-1}) + C(\xi_t, y_t|\mathcal{I}_{t-1})D^{-1}(y_t|\mathcal{I}_{t-1})\{y_t - E(y_t|\mathcal{I}_{t-1})\} \\ &= x_{t|t-1} + P_{t|t-1}H_t'F_t^{-1}e_t. \end{aligned} \tag{189}$$

The dispersion matrix under (178) for the updated estimate is obtained via equation (180):

$$\begin{aligned} D(\xi_t|\mathcal{I}_t) &= D(\xi_t|\mathcal{I}_{t-1}) - C(\xi_t, y_t|\mathcal{I}_{t-1})D^{-1}(y_t|\mathcal{I}_{t-1})C(y_t, \xi_t|\mathcal{I}_{t-1}) \\ &= P_{t|t-1} - P_{t|t-1}H_t'F_t^{-1}H_tP_{t|t-1}. \end{aligned} \tag{190}$$

The set of information $\mathcal{I}_t = \{y_t, \ldots, y_1, \mathcal{I}_t\}$, on which the Kalman filter estimates are based, can be represented, equivalently, by replacing the sequence $\{y_t, \ldots, y_1\}$ of observations by the sequence $\{e_t, \ldots, e_1\}$ of the prediction errors, which are mutually uncorrelated.

The equivalence can be demonstrated by showing that, given the initial information of $\mathcal{I}_0$, there is a one-to-one correspondence between the two sequences, which depends only on the known parameters of equations (169), (170) and (171). The result is intuitively intelligible, for, at each instant t, the prediction error e_t contains only the additional information of y_t that is not predictable from the information in the set $\mathcal{I}_{t-1}$; which is to say that $\mathcal{I}_t = \{e_t, \mathcal{I}_{t-1}\}$.

The prediction errors provide a useful formulation of the likelihood function from which the parameters that are assumed to be know to the Kalman filter can be estimated from the data. Under the assumption that the disturbances are normally distributed, the likelihood function is given by

$$\ln L = -\frac{kT}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln|F_t| - \frac{1}{2}\sum_{t=1}^{T}e_t'F_t^{-1}e_t. \tag{191}$$

This form was proposed originally by Schweppe (1965). It tractability, which is a partial compensation for the complexity of the Kalman filter, has contributed significantly to the popularity of the state-space formulation of the structural time-series models.

There are various ways in which the value of the initial condition in $\mathcal{I}_0 = \{\xi_0, P_0\}$ may be obtained. If the processes are stationary, then the eigenvalues of the transition matrix Φ must lie within unit circle, which implies that $\lim(n \to \infty)\Phi^n = 0$. Then, there is $E(\xi_0) = x_0 = 0$ and $D(\xi_0) = P_0 = \Phi P_0 \Phi' + \Psi$; and the latter equation may be solved by analytic or iterative means for the value of P_0.

In the nonstationary case, the initial conditions require to be determined in the light of the data. To allow the information of the data rapidly to assert itself, one may set $P_0 = \lambda I$, where λ is given a large value. This will associate a large dispersion to the initial state estimate x_0 to signify a lack of confidence in its value, which will allow the estimate to be enhanced rapidly by the information of the data points. Using the terminology of Bayesian estimation, this recourse may be described as the method of the diffuse prior.

Data-dependent methods for initialising the Kalman filter of a more sophisticated nature, which make amends for, or which circumvent, the arbitrary choices of x_0 and P_0, have been proposed by Ansley and Kohn (1982) and by de Jong (1991), amongst others. These methods have been surveyed by Pollock (2003). Another account of the method of Ansley and Kohn, which is more accessible than the original one, has also been provided by Durbin and Koopman (2001).

The method of the diffuse prior bequeaths some pseudo information to the Kalman filter, in the form of arbitrary initial conditions, which remains in the system indefinitely, albeit that its significance is reduced as the sample information is accumulated. The technique of Ansley and Kohn is designed to remove the pseudo information at the earliest opportunity, which is when there is enough sample information to support the estimation of the state vector.

In their exposition of the technique, Ansley and Kohn described a transformation of the likelihood function that would eliminate its dependence on the initial conditions. This transformation was a purely theoretical device without any practical implementation. However, it is notable that the method of handling the start-up problem that has been expounded in section 8, which employs a differencing operation to reduce the data sequence to stationarity, has exactly the effect of eliminating the dependence upon initial conditions.

11.1 The Smoothing Algorithms

The Kalman filter generates an estimate $x_t = E(\xi_t|\mathcal{I}_t)$ of the current state of the system using information from the past and the present. To derive a more efficient estimate, we should take account of information that arises subsequently up to the end of the sample. Such an estimate, which may be denoted by $x_{t|T} = E(\xi_t|\mathcal{I}_T)$, is described as a fixed-interval estimate; and the various algorithms that provide the estimate are described as a fixed-interval smoothers.

It is laborious to derive the smoothing algorithms, of which there exist a fair variety. The matter is treated at length in the survey article of Merkus, Pollock and de Vos (1993) and in the monograph of Weinert (2001). Econometricians and others have derived a collection of algorithms which are, in some respects, more efficient in computation than the classical fixed-interval smoothing algorithm that is due to Rauch (1963), of which a derivation can be found in Anderson and Moore (1979), amongst other sources. A variant of the classical algorithm has been employed by Young *et al.* (2004) in the CAPTAIN MatLab toolbox, which provides facilities for estimating structural time-series models.

The classical algorithm may be derived via a sleight of hand. Consider enhancing the estimate $x_t = E(\xi_t|\mathcal{I}_t)$ in the light of the information afforded by an exact knowledge of the subsequent state vector ξ_{t+1}. The information would be conveyed by

$$h_{t+1} = \xi_{t+1} - E(\xi_{t+1}|\mathcal{I}_t), \tag{192}$$

which would enable us to find

$$E(\xi_t|\mathcal{I}_t, h_{t+1}) = E(\xi_t|\mathcal{I}_t) + C(\xi_t, h_{t+1}|\mathcal{I}_t)D^{-1}(h_{t+1}|\mathcal{I}_t)h_{t+1}. \tag{193}$$

Here there are

$$\begin{aligned} C(\xi_t, h_{t+1}|\mathcal{I}_t) &= E\{\xi_t(\xi_t - x_t)'\Phi' + \xi_t\nu_t'\big|\mathcal{I}_t\} = P_t\Phi' \quad \text{and} \\ D(h_{t+1}|\mathcal{I}_t) &= P_{t+1|t}. \end{aligned} \tag{194}$$

It follows that

$$E(\xi_t|\mathcal{I}_t, h_{t+1}) = E(\xi_t|\mathcal{I}_t) + P_t\Phi'P_{t+1|t}^{-1}\Big\{\xi_{t+1} - E(\xi_{t+1}|\mathcal{I}_t)\Big\}. \tag{195}$$

Of course, the value of ξ_{t+1} in the RHS of this equation is not observable. However, if we take the expectation of the equation conditional upon the available information of the set $\mathcal{I}_T$, then ξ_{t+1} is replaced by $E(\xi_{t+1}|\mathcal{I}_T)$ and we get a formula that can be rendered as

$$x_{t|T} = x_t + P_t\Phi'P_{t+1|t}^{-1}\{x_{t+1|T} - x_{t+1|t}\}. \tag{196}$$

The dispersion of the estimate is given by

$$P_{t|T} = P_t - P_t\Phi'P_{t+1|t}^{-1}\{P_{t+1|t} - P_{t+1|T}\}P_{t+1|t}^{-1}\Phi P_t. \tag{197}$$

This derivation was published by Ansley and Kohn (1982). It highlights the notion that the information that is used in enhancing the estimate of ξ_t is contained entirely within the smoothed estimate of ξ_{t+1}.

The smoothing algorithm runs backwards through the sequence of estimates generated by the Kalman filter, using a first-order feedback in respect of the smoothed estimates. The estimate $x_t = E(\xi_t|\mathcal{I}_t)$ is enhanced in the light of the "prediction error"$x_{t+1|T} - x_{t+1|t}$, which is the difference between the smoothed and the unsmoothed estimates of the state vector ξ_{t+1}.

In circumstances where the factor $P_t\Phi'P_{t+1|t}^{-1}$ can be represented by a constant matrix, the classical algorithm is efficient and easy to implement. This would be the case if there were a constant transition matrix Φ and if the filter gain K_t had converged to a constant. In all other circumstances, where it is required recompute the factor at each iteration of the index t, the algorithm is liable to cost time and to invite numerical inaccuracies. The problem, which lies with the inversion of $P_{t+1|t}$, can be avoided at the expense of generating a supplementary sequence to accompany the smoothing process.

11.2 Equivalent and Alternative Procedures

The derivations of the Kalman filter and the fixed-interval smoothing algorithm are both predicated upon the minimum-mean-square-error estimation criterion. Therefore, when the filter is joined with the smoothing algorithm, the resulting estimates of the data components should satisfy this criterion. However, its fulfilment will also depend upon an appropriate choice of the initial conditions for the filter. For this, one may use the method of Ansley and Kohn (1985).

The same criterion of minimum-mean-square-error estimation underlies the derivation of the finite-sample Wiener–Kolmogorov filter that has been presented in sections 7 and 8. Therefore, when they are applied to a common model, the Wiener–Kolmogorov filter and the combined Kalman filter and smoother are expected to deliver the same estimates.

The handling of the initial-value problem does appear to be simpler in the Wiener–Kolmogorov method than in the method of Ansley and Kohn. However, the finite-sample Wiener–Kolmogorov method of section 8 is an instance of the transformation approach that Ansley and Kohn have shown to be equivalent to their method.

It should be noted that the minimum-mean-square-error estimates can also be obtained using a time-invariant version of the Wiener–Kolmogorov filter, provided that the finite data sequence can be extended by estimates of the presample and post-sample elements. However, this requires that the filter should relate to a well-specified ARMA or ARIMA model that is capable of generating the requisite forecasts and backcasts. If this is the case, then a cogent procedure for generating the extra-sample elements is the one that has been been described by Burman (1980) and which is incorporated in the TRAMO–SEATS program.

The upshot is that several routes lead to the same ends, any of which may be taken. Nevertheless, there have been some heated debates amongst econometrics who are the proponents of alternative approaches. However, the only significant issue is the practical relevance of the alternative models that are intended to represent the processes that underlie the data or to provide heuristic devices for generating the relevant filters.

An agnostic stance has been adopted in this chapter; and no firm pronouncements have been made concerning the nature of economic realities. Nevertheless, it has been proposed that the concept of a band-limited process, which had been largely overlooked in the past, is particularly relevant to this area of econometric analysis.

This concept encourages consideration of the Fourier methods of filtering of section 9, which are capable of separating components of the data that lie in closely adjacent frequency bands, as is the case in Figure 10, where the fundamental seasonal component abuts the low-frequency structure of the trend-cycle component. Such methods have been explored in greater detail in a paper of Pollock (2008); and they have been implemented in a program that is available from a website at the address

`http://www.le.ac.uk/users/dsgp1/`

References

Anderson BDO and Moor JB 1979 *Optimal Filtering.* Prentice–Hall, Englewood Cliffs, New Jersey.

Ansley CF and Kohn R 1982 A geometrical derivation of the fixed interval smoothing equations. *Biometrika* **69**, 486–7.

Ansley CF and R Kohn 1985 Estimation, Filtering and Smoothing in State Space Models with Incompletely Specified Initial Conditions. *The Annals of Statistics* **13**, 1286–1316.

Banerjee A Dolado J Galbraith JW and Hendry DF 1993 *Co-integration, Error-correction and the Econometric Analysis of Non-stationary Data.* Oxford University Press, Oxford.

Baxter M and King RG 1999 Measuring business Ccycles: approximate bandpass filters for economic time series. *Review of Economics and Statistics* **81**, 575–593.

Bell W 1984 Signal extraction for nonstationary time series. *The Annals of Statistics* **12**, 646–664.

Burman JP 1980 Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society, Series A* **143**, 321–337.

Caporello G and Maravall A 2004 *Program TSW: Revised Reference Manual.* Working Paper, Research Department, Banco de España.

Dagum EB and Luati A 2004 A linear transformation and its properties with special applications in time series filtering. *Linear Algebra and its Applications* **338**, 107–117.

Daubechies I 1992 *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, Philadelphia.

de Jong P 1991 The diffuse Kalman filter. *The Annals of Statistics* **19**, 1073–1083.

Dirac PAM 1958 *The Principles of Quantum Mechanics, Fourth Edition.* Oxford University Press, Oxford.

Doherty M 2001 The surrogate Henderson filters in X-11. *Australian and New Zealand Journal of Statistics* **43**, 385–392.

Durbin J and Koopman SJ 2001 *Time Series Analysis by State Space Methods.* Oxford University Press.

Eurostat, 2002 *Demetra 2.0 User Manual: Seasonal Adjustment Interface for TRAMO–SEATS and X12-ARIMA,* The Statistical Office of the European Communities.
`http://circa.europa.eu/irc/dsis/eurosam/info/data/demetra.htm.`

Fan J and Gijbels I 1996 *Local Polynomial Modelling and Its Applications.* (Monographs on Statistics and Applied Probability) Chapman and Hall, London.

Godolphin EJ 1976 On the Cramér–Wold factorisation. *Biometrika* **63**, 367–380.

Godolphin, EJ and Unwin JM 1983 Evaluation of the covariance matrix for the maximum likelihood estimator of a Gaussian autoregressive-moving average process. *Biometrika* **70**, 279–284.

Gómez V and Maravall A 1996 *Programs TRAMO and SEATS: Instructions for the user, (with some updates).* Working Paper 9628, Servicio de Estudios, Banco de España.

Gray AG and Thomson PJ 2002 On a family of finite moving-average trend filters for the ends of series. *Journal of Forecasting* **21**, 125–149.

Harrison PJ and Stevens CF 1971 A Bayesian approach to short-term forecasting. *Operational Research Quarterly* **22**, 341–362.

Harrison PJ and Stevens CF 1976 Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Series B* **38**, 205–247.

Harvey AC 1989 *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

Henderson R 1916 Note on graduation by adjusted average. *Transactions of the Actuarial Society of America* **17**, 43–48.

Henderson R 1924 A new method of graduation. *Transactions of the Actuarial Society of America* **25**, 29–40.

Higgins JR 1985 Five short stories about the cardinal series. *Bulletin of the American Mathematical Society* (N.S.) **12**, 45–89.

Hillmer SC and Tiao GC 1982 An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association* **77**, 63–70.

Kaiser Regina and Maravall A 2001 *Measuring Business Cycles in Economic Time Series.* Springer Lecture Notes in Statistics, Springer-Verlag, New York.

Kolmogorov A N Stationary Sequences in a Hilbert Space, (in Russian), *Bulletin of the Moscow State University,* **2**, 1–40.

Koopman SJ, Harvey AC, Doornik JA and Shephard N 2000 *STAMP 6.0: Structural Time Series Analyser, Modeller and Predictor.* Timberlake Consultants Press, London.

Kenny PB and Durbin J 1982 Local trend estimation and seasonal adjustment of economic and social time series. *Journal of the Royal Statistical Society, Series A* **145**, 1–41.

Kitagawa G and Gersch W 1996 *Smoothness Priors Analysis of Time Series.* Springer-Verlag, New York.

Kolmogorov AN 1941 Interpolation and extrapolation. *Bulletin de l'Academie des Sciences de U.S.S.R* Ser. Math **5**, 3–14.

Luke HD 1999 The origins of the sampling theorem. *IEEE Communications Magazine* **37**, 106–108.

Maravall A and Pierce DA 1987 A prototypical seasonal adjustment model. *Journal of Time Series Analysis* **8,** 177–193.

McElroy T 2006 *Matrix Formulas for Nonstationary ARIMA Signal Extraction.* U.S. Census Bureau.

Merkus HR, Pollock DSG and de Vos AF 1993 A synopsis of the smoothing formulae associated with the Kalman filter. *Computational Economics* **6**, 177–200.

Monsell BC, Aston JAD and Koopman SJ 2003 *Toward X-13?* U. S. Census Bureau, Washington, DC, USA. (`http://www.census.gov/srd/www/sapaper.html`)

Musgrave JC 1964a *A Set of End Weights to End all End Weights.* Unpublished Working Paper of the U.S. Bureau of Commerce.
(`http://www.census.gov/ts/papers/Musgrave1964a.pdf`)

Musgrave JC 1964b Alternative Sets of Weights Proposed for X-11 Seasonal Factor Curve Moving Averages, Unpublished Working Paper of the U.S. Bureau of Commerce.
(`http://www.census.gov/ts/papers/Musgrave1964a.pdf`)

Percival DB and Walden AT 2000 *Wavelet Methods for Time Series Analysis.* Cambridge University Press, Cambridge.

Pollock DSG 1999 *A Handbook of Time-Series Analysis, Signal Processing and Dynamics.* Academic Press, London.

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

Pollock DSG 2002 A review of TSW: the Windows version of the TRAMO-SEATS program. *Journal of Applied Econometrics* **17**,291–299.

Pollock DSG 2003 Recursive estimation in econometrics. *Journal of Computational Statistics and Data Analysis* **44**, 37–75.

Pollock DSG and Lo Cascio Iolanda 2006 Non-Dyadic Wavelet Analysis. In *Advances in Computational Economics Finance and Management Science.* Springer Verlag.

Pollock DSG 2008 Realisations of finite-sample frequency-selective filters. *Journal of Statistical Planning and Inference* doi: 10.1016/j.jspi.2008.08.020.

Proietti T 2002 Forecasting with Structural Time Series Models. In *A Companion to Economic Forecasting,* (ed. Clements M). Blackwell Publishers, Oxford.

Proietti T and Luati Alessandra 2006 Least squares regression: graduation and filters. In *Measurement in Economics: A Handbook* (ed. Boumans M). Academic Pess.

Rauch HE 1963 Solutions to the linear smoothing problem. *IEEE Transactions on Automatic Control* **AC-8**, 371–372.

Schweppe FC 1965 Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory,* **11**, 61–70.

Schiff LI 1981 *Quantum Mechanics.* McGraw-Hill Book Co.

Shiskin J, Young AH and Musgrave JC 1967 *The X-11 Variant of the Census Method II Seasonal Adjustment Program.* Technical Paper 15, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Simonoff JS 1996 *Smoothing Methods in Statistics.* Springer Series in Statistics Springer, New York.

Wallis KF 1981 Models for X-11 and X-11-Forecast Procedures for Preliminary and Revised Seasonal Adjustments. *Proceedings of the Conference on Applied Time Series Analysis of Economic Data*, A.S.A–Census–NBER, pp. 3–11.

Weinert HL 2001 *Fixed Interval Smoothing for State Space Models.* Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, Dordrecht.

West M 1997 Time series decomposition. *Biometrika* **84**, 489–494.

West M and Harrison J 1997 *Bayesian Forecasting and Dynamic Models, 2nd edition.* Springer-Verlag, New York.

West M, Prado R and Krystal AD 1999 Evaluation and comparison of EEG traces: latent structure in nonstationary time series. *Journal of the American Statistical Association* **94**, 1083–1095.

Whittaker ET 1915 On functions that are represented by expansions of the interpolation theory. *Proceedings of the Royal Society of Edinburgh* **35**, 181–194.

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

Wiener N 1941 *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* Report on the Services Research Project DIC-6037. Published in book form in 1949 by MIT Technology Press and John Wiley and Sons, New York.

Wilson G 1969 Factorisation of the covariance generating function of a pure moving average process. *SIAM Journal of Numerical Analysis* **6**, 1–7.

Wold H 1954 *A Study in the Analysis of Stationary Time Series,* 2nd edition. Almquist and Wiksell, Stockholm.

Young PC, Taylor CJ, Tych W, Pedregal DJ and McKenna CJ 2004 The Captain Toolbox, Centre for Research on Environmental Systems and Statistics, Lancaster University. (`www.es.lancs.ac.uk/cres/captain`).