LECTURE 9

Filtering and Trend Extraction

Trends and Unit Root Processes

Following the persuasive advocacy by Box and Jenkins in the 1970's, the difference operator has assumed a central role in the econometric analysis of trended data sequences. Box and Jenkins (1972) proposed to use the difference operator in conjunction with ordinary autoregressive moving-average (ARMA) models. When an appropriate power of the difference operator is embedded in the autoregressive polynomial, the effect is to create an autoregressive integrated moving-average (ARIMA) model, which can generate a wide variety of sequences that resemble those that are encountered in econometric analyses.

An advantage of the ARIMA formulation is that it enables the methodology that is appropriate for identifying and estimating stationary ARMA processes to be extended, with no extra complication, to accommodate nonstationary processes. The difference operator that is embedded in the autoregressive polynomial can be applied to the trended data sequence in order to reduce it to stationarity. Thereafter, the data can be modelled via an ordinary ARMA process. When the ARMA process has been identified and estimated, it can converted to an ARIMA process by applying the summation operator, which is the reverse of the difference operator.

The difference operator has a powerful effect upon the data. It nullifies the trend and it severely attenuates the elements of the data that are adjacent in frequency to the zero frequency of the trend. It also amplifies the high-frequency elements of the data. The effect is apparent in Figure 1, which shows the squared gain of the difference operator. The figure also shows the squared gain of the summation operator, which gives unbounded power to the elements that have frequencies in the vicinity of zero.

The relevant information of an econometric sequence often lies in the low-frequency region, and is it difficult to discern it clearly within a differenced sequence. The effects of the difference operator can be mitigated within an

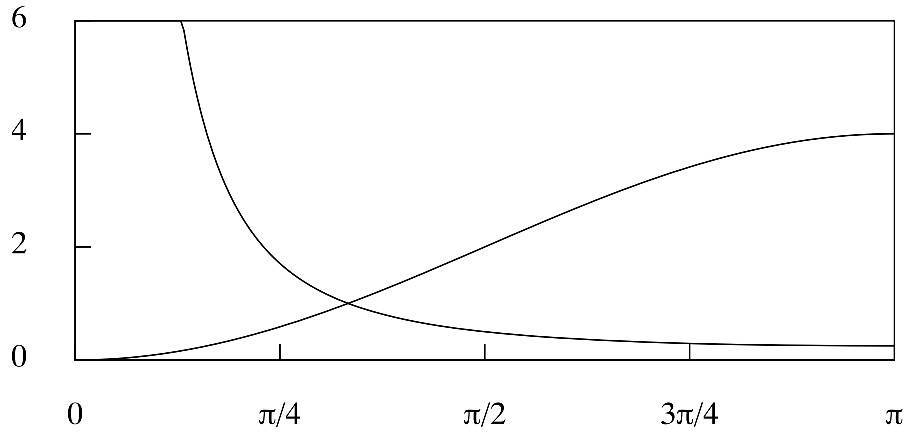

Figure 1. The squared gain of the difference operator, which has a zero at zero frequency, and the squared gain of the summation operator, which is unbounded at zero frequency.

ARIMA model by the roots of the moving-average operator, which can counteract the unit roots of the autoregressive difference operator. However, to use the difference operator in isolation to reduce the data to stationarity is a drastic recourse, which has been likened to throwing the baby out with the bath water. Therefore, we must seek other methods of detrending the data.

The simplest of the so-called unit-root processes that incorporate an autoregressive difference operator is the ordinary random walk. This is often represented in terms of the lag operator L by the equation

$$(1 - L)y(t) = \varepsilon(t), \tag{1}$$

where $\varepsilon(t) = \{\varepsilon_0, \pm\varepsilon_1, \pm\varepsilon_2, \ldots\}$ denotes a doubly-infinite mean-zero sequence of independently and identically distributed random variables, described as a white-noise sequence. The inverse of the difference operator is the summation operator

$$\Sigma = (1 - L)^{-1} = (1 + L + L^2 + \cdots). \tag{2}$$

Multiplying equation (1) by Σ gives

$$\begin{aligned} y(t) &= \Sigma\varepsilon(t) = (1 + L + L^2 + \cdots)\varepsilon(t) \\ &= \{\varepsilon(t) + \varepsilon(t-1) + \varepsilon(t-2) + \cdots\} \end{aligned} \tag{3}$$

The generic element of $y(t)$ is a sum of an infinite number of lagged variables. Since these are independently and identically distributed with a finite variance, the elements of $y(t)$ must have an infinite variance. Therefore, there is a zero probability that any element will fall within a finite distance of zero. This is a cause for some embarrassment, for it implies that equations (1) and (3) are inappropriate descriptions of any nonstationary data sequence that has elements that are bounded in value.

The squared gain of the summation operator that is represented in Figure 1 can also be interpreted as the pseudo spectral density function of a random walk driven by a white-noise process of variance 2π. This random walk, which is defined over a doubly-infinite set of positive and negative integers, is a theoretical process of doubtful reality.

The appropriate representation of a random walk for present purposes is one that defines a finite starting value at a definite date and which incorporates the corresponding initial conditions. We shall consider this matter in the following section, which describes alternative representations of ARMA and ARIMA processes.

Representations of ARMA and ARIMA Processes

An autoregressive moving-average (ARMA) model can be represented by the equation

$$\sum_{i=0}^{p} \phi_i y_{t-i} = \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}, \quad \text{with} \quad \phi_0 = \theta_0 = 1. \tag{4}$$

The normalisation of ϕ_0 indicates that y_t is the dependent variable of this equation. By setting $t = 0, 1, \ldots, T-1$, a set of T equations is generated that can be arrayed in a matrix format as follows:

$$\begin{bmatrix} y_0 & y_{-1} & \cdots & y_{-p} \\ y_1 & y_0 & \cdots & y_{1-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_p & y_{p-1} & \cdots & y_0 \\ \vdots & \vdots & & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-p-1} \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \varepsilon_0 & \varepsilon_{-1} & \cdots & \varepsilon_{-q} \\ \varepsilon_1 & \varepsilon_0 & \cdots & \varepsilon_{1-q} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_q & \varepsilon_{q-1} & \cdots & \varepsilon_0 \\ \vdots & \vdots & & \vdots \\ \varepsilon_{T-1} & \varepsilon_{T-2} & \cdots & \varepsilon_{T-q-1} \end{bmatrix} \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_q \end{bmatrix}. \tag{5}$$

Apart from the elements $y_0, y_1, \ldots, y_{T-1}$ and $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{T-1}$, which fall within the indicated sample period, these equations comprise the values $y_{-p}, \ldots, y_{-1}$ and $\varepsilon_{-q}, \ldots, \varepsilon_{-1}$, which are to be found in the top-right corners of the matrices, and which constitute the initial conditions at the start-up time of $t = 0$. The presence of these initial conditions allows this format to be used both for stationary ARMA processes and for nontstationary ARIMA processes.

Each of the elements within the display of (5) can be associated with the power of z that is indicated by the value of its subscripted index. In that case, the equations of the system become

$$\sum_{i=0}^{p}\{\phi_i z^i\}\{y_{t-i}z^{t-i}\} = \sum_{i=0}^{q}\{\theta_i z^i\}\{\varepsilon_{t-i}z^{t-i}\} \tag{6}$$

or, equivalently,

$$z^t\sum_{i=0}^{p}\phi_i y_{t-i} = z^t\sum_{i=0}^{q}\theta_i\varepsilon_{t-i}.$$

A comparison can be made between the equations above and the matrix equations that correspond to the z-transform polynomial equation

$$\phi(z)y(z) = \theta(z)\varepsilon(z), \tag{7}$$

wherein

$$\begin{aligned}\phi(z) &= 1+\phi_1 z+\cdots+\phi_p z^p \quad \text{and}\\ \theta(z) &= 1+\theta_1 z+\cdots+\theta_q z^q.\end{aligned} \tag{8}$$

Often, the data sequences are taken to be doubly infinite, which gives rise to $y(z) = \{y_0 \pm y_1 z \pm y_2 z^2 \pm \ldots\}$ and $\varepsilon(z) = \{\varepsilon_0 \pm \varepsilon_1 z \pm \varepsilon_2 z^2 \pm \ldots\}$. Consider, instead, the z-transforms of the finite sequences running from $t = 0$ to $t = T-1$:

$$\begin{aligned}y(z) &= y_0 + y_1 z + \cdots + y_{T-1}z^{T-1},\\ \varepsilon(z) &= \varepsilon_0 + \varepsilon_1 z + \cdots + \varepsilon_{T-1}z^{T-1}.\end{aligned} \tag{9}$$

Then, the corresponding matrix expression is

$$\begin{bmatrix} y_0 & 0 & \ldots & 0\\ y_1 & y_0 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ y_p & y_{p-1} & \ldots & y_0\\ \vdots & \vdots & & \vdots\\ y_{T-1} & y_{T-2} & \ldots & y_{T-p-1}\\ 0 & y_{T-1} & \ldots & y_{T-p}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & y_{T-1}\end{bmatrix}\begin{bmatrix}1\\ \phi_1\\ \vdots\\ \phi_p\end{bmatrix} = \begin{bmatrix} \varepsilon_0 & 0 & \ldots & 0\\ \varepsilon_1 & \varepsilon_0 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ \varepsilon_q & \varepsilon_{q-1} & \ldots & \varepsilon_0\\ \vdots & \vdots & & \vdots\\ \varepsilon_{T-1} & \varepsilon_{T-2} & \ldots & \varepsilon_{T-q-1}\\ 0 & \varepsilon_{T-1} & \ldots & \varepsilon_{T-q}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \varepsilon_{T-1}\end{bmatrix}\begin{bmatrix}1\\ \theta_1\\ \vdots\\ \theta_q\end{bmatrix}. \tag{10}$$

This system is afflicted both by the absence of the appropriate initial conditions, indicated by the zeros in the upper right corners of the matrices, and by the presence of end effects, corresponding to the upper-triangular matrices appended at the bottom of both arrays.

Notice, however, that, under any circumstances, the equations that are associated with $z^r, \ldots, z^{T-1}$, where $r = \max(p, q)$, will provide a valid description of the corresponding data segment $\{y_r, \ldots, y_{T-1}\}$. The equations can be denoted by $\phi(z)y(z) = \theta(z)\varepsilon(z) - Q(z)$, where $Q(z)$ is a polynomial in the various powers of z that are associated with the unwanted end-effects.

With $Q(z)$ included, it will be allowable freely to manipulate the polynomial algebra associated with the finite realisation of the ARMA process, while paying little or no attention to the end effects. In fact, we shall hereafter omit to include this term within the algebra on the understanding that either it is invisibly present or else that our attention is confined to the appropriate range of the temporal index, which is $t = r, \ldots, T-1$

In certain circumstances, where the data are stationary and the processes that generate them are reasonably constant, it is possible to overcome both effects by treating the data as if they were generated by a circular process. By drafting the triangular arrays at the bottom of the matrices into the spaces at their tops that are occupied by zeros, the following system is created:

$$\begin{bmatrix} y_0 & y_{T-1} & \cdots & y_{T-p} \\ y_1 & y_0 & \cdots & y_{T+1-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_p & y_{p-1} & \cdots & y_0 \\ \vdots & \vdots & & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-p-1} \end{bmatrix} \begin{bmatrix} 1 \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \varepsilon_0 & \varepsilon_{T-1} & \cdots & \varepsilon_{T-q} \\ \varepsilon_1 & \varepsilon_0 & \cdots & \varepsilon_{T+1-q} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_q & \varepsilon_{q-1} & \cdots & \varepsilon_0 \\ \vdots & \vdots & & \vdots \\ \varepsilon_{T-1} & \varepsilon_{T-2} & \cdots & \varepsilon_{T-q-1} \end{bmatrix} \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_q \end{bmatrix}. \tag{11}$$

We might continue to associate powers of z with the elements of this system in accordance with the values of their subscripts. However, the powers of some of the arguments that can be put in place of z have a modulus T, which is to say that $z^{T-t} = z^{-t}$. These are the so-called circular arguments. In such cases, the system of (11) will correspond to the polynomial equation of (7) as well as representing an instance of the system of (5).

A leading example of a circular argument is the complex exponential $z = \exp\{i2\pi j/T\}$, wherein $j = 0, 1, \ldots, T-1$. Use of this argument carries the analysis into the frequency domain of time-series analysis. Another example of a circular argument, which keeps the analysis within the time domain, is the circulant matrix

$$K_T = [\, e_1 \quad e_2 \quad \ldots \quad e_{T-2} \quad e_0 \,] = \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}. \tag{12}$$

This is obtained from the identity matrix $I_T = [e_0, e_1, \ldots, e_{T-1}]$ by transposing the leading vector e_0 to the end of the array. The powers $K_T^0 = I_T,\ K_T,\ K_T^2,\ \ldots,\ K^{T-1}$ form a cycle and, moreover, $K^{T+r} = K^r$. Setting $z = K_T$ within the polynomial z-transform equation of (7) generates the matrix expression of (11).

An alternative replacement for z within the (7) is the matrix lag operator $L_T = [e_1, e_2, \ldots, e_{T-2}, 0]$ which differs from the circulant matrix K_T by having a column of zeros in place of the final column e_0. It will be seen that, for $r \in \{0, 1, \ldots, T-1\}$, the matrix L^r has units on the r-th sub diagonal band and for zeros elsewhere, whereas, for $r \geq T$, there is $L^r = 0$.

It follows that, by putting $z = L_T$ in equation (7), a matrix system is generated that corresponds to the system obtained from (10) by removing the triangular end-effects from the bottom of the matrices. This gives a version of equation (5) in which the pre sample values are set to zeros.

The Integration of ARMA Process

The conversion of an ARMA process to an integrated ARIMA process is a straightforward matter of summing its elements. A d-fold summation will turn an ARMA(p, q) into an ARIMA(p, d, q) process.

The summation operator is the inverse of the difference operator. The z-transform of the d-fold summation operator is as follows:

$$\Sigma^d(z) = \frac{1}{(1-z)^d} = 1 + dz + \frac{d(d+1)}{2!} + \frac{d(d+1)(d+2)}{3!} + \cdots . \tag{13}$$

The coefficients of the negative binomial expansion are readily calculated via the repeated summation of a sequence of units. The following three matrices, which illustrate the fact, are generated by replacing z within $\Sigma(z)$, $\Sigma^2(z)$ and $\Sigma^3(z)$, respectively, by the matrix lag operator L_4:

(14)

$$\Sigma_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \Sigma_4^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ 4 & 3 & 2 & 1 \end{bmatrix}, \quad \Sigma_4^3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 6 & 3 & 1 & 0 \\ 10 & 6 & 3 & 1 \end{bmatrix}.$$

Consider the ARIMA(p, d, q) process described by the equation

$$\nabla^d(z)\alpha(z)y(z) = \alpha(z)g(z) = \theta(z)\varepsilon(z), \tag{15}$$

where $g(t) = \nabla^d(z)y(t)$ denotes the ARMA process that is obtained by d-fold differencing. Then, the manner in which the integrated process is obtained from the stationary process is described by the equation $y(t) = \nabla^{-d}(z)g(z) = \Sigma^d(z)g(z)$.

In the case where $d = 1$ and where $g(z)$ comprises T observations, there is a single integration, which is effected by the matrix Σ_T. Then, each element of the ARMA process gives rise of a constant function, which is to say that its effect will persist throughout the sample. In the case of $d = 2$, each element will give rise to a linear trend and, in the case of $d = 3$, each will give rise of a quadratic trend.

In the so-called Beveridge–Nelson decomposition, the unit roots of the autoregressive operator are separated from the stable roots of by a partial-fraction decomposition. Assuming that the degree of the numerator polynomial $\theta(z)$ is less than that of the denominator polynomial $\nabla^d(z)\alpha(z)$, the decomposition of $y(z)$ takes the form of

$$y(t) = \frac{\theta(z)}{\nabla^d(z)\alpha(z)}\varepsilon(z) = \frac{\kappa(z)}{\nabla^d(z)}\varepsilon(z) + \frac{\lambda(z)}{\alpha(z)}\varepsilon(z). \tag{16}$$

Thus, $y(t)$ is expressed as the sum of an integrated process and a stationary process. Both processes are driven by the same stationary white-noise process, which is represented by $\varepsilon(z)$.

On the assumption that the parameters of the ARMA process are known, the polynomial $\varepsilon(z)$ would be fully recoverable from $y(z)$; and then the non stationary component of the decomposition would be given by

$$x(z) = \left\{ \frac{\kappa(z)}{\nabla^d(z)} \times \frac{\nabla^d(z)\alpha(z)}{\theta(z)} \right\} y(z) = \frac{\kappa(z)\alpha(z)}{\theta(z)} y(z). \tag{17}$$

In practice, the estimates of the two components will be derived from the residual sequence obtained in estimating the parameters of the ARMA process described by the equation $\alpha(z)g(z) = \theta(z)\varepsilon(z)$.

Polynomial Interpolation

The first p columns of the matrix Σ_T^p provide a basis of the set of polynomials of degree $p-1$ defined on the set of integers $t = 0, 1, 2, \ldots, T-1$. An example is provided by the first three columns of the matrix Σ_4^3, which may be transformed as follows:

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 6 & 3 & 1 \\ 10 & 6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -2 & -1 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}. \tag{18}$$

The first column of the matrix on the LHS contains the ordinates of the quadratic function $(t^2 + t)/2$. The columns of the transformed matrix are recognisably the ordinates of the powers t^0, t^1 and t^2 corresponding to the integers $t = 1, 2, 3, 4$. The natural extension of the matrix to T rows provides a basis for the quadratic functions $q(t) = at^2 + bt + c$ defined on T consecutive integers.

The matrix of the powers of the integers is notoriously ill-conditioned. In calculating polynomial regressions of any degree in excess of the cubic, it is advisable to employ a basis of orthogonal polynomials, for which purpose some specialised numerical procedures are available. However, in the present context, which concerns the differencing and the summation of econometric data sequences, the degree in question rarely exceeds two. Nevertheless, it is appropriate to consider the algebra of the general case.

Consider, therefore, the matrix that takes the p-th difference of a vector of order T, which is

$$\nabla_T^p = (I - L_T)^p. \tag{19}$$

This matrix can be partitioned so that $\nabla_T^p = [Q_*, Q]'$, where Q_*' has p rows. If y is a vector of T elements, then

$$\nabla_T^p y = \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} y = \begin{bmatrix} g_* \\ g \end{bmatrix}; \tag{20}$$

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

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

$$\begin{bmatrix} S_* & S \end{bmatrix} \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} = S_* Q_*' + SQ' = I_T, \tag{21}$$

and that

$$\begin{bmatrix} Q_*' \\ Q' \end{bmatrix} \begin{bmatrix} S_* & S \end{bmatrix} = \begin{bmatrix} Q_*' S_* & Q_*' S \\ Q' S_* & Q' S \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_{T-p} \end{bmatrix}. \tag{22}$$

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

$$y = S_* g_* + Sg. \tag{23}$$

Since the submatrix S_*, provides a basis for all polynomials of degree $p-1$ that are defined on the integer points $t = 0, 1, \ldots, T-1$, it follows that $S_* g_* = S_* Q_*' y$ contains the ordinates of a polynomial of degree $p-1$, which is interpolated through the first p elements of y, indexed by $t = 0, 1, \ldots, p-1$, and which is extrapolated over the remaining integers $t = p, p+1, \ldots, T-1$.

A polynomial that is designed to fit the data should take account of all of the observations in y. Imagine, therefore, that $y = \phi + \eta$, where ϕ contains the ordinates of a polynomial of degree $p-1$ and η is a disturbance term with $E(\eta)=0$ and $D(\eta)=\sigma_\eta^2 I_T$. Then, in forming an estimate $x=S_*r_*$ of ϕ, we should minimise the sum of squares $\eta'\eta$. Since the polynomial is fully determined by the elements of a starting-value vector r_*, this is a matter of minimising

$$(y-x)'(y-x)=(y-S_*r_*)'(y-S_*r_*) \tag{24}$$

with respect to r_*. The resulting values are

$$r_*=(S_*'S_*)^{-1}S_*'y \quad \text{and} \quad x=S_*(S_*'S_*)^{-1}S_*'y. \tag{25}$$

An alternative representation of the estimated polynomial is available. This is provided by the identity

$$S_*(S_*'S_*)^{-1}S_*'=I-Q(Q'Q)^{-1}Q'. \tag{26}$$

To prove this identity, consider the fact $Z=[Q,S_*]$ is square matrix of full rank and that Q and S_* are mutually orthogonal such that $Q'S_*=0$. Then

$$\begin{aligned} Z(Z'Z)^{-1}Z' &= \begin{bmatrix} Q & S_* \end{bmatrix} \begin{bmatrix} (Q'Q)^{-1} & 0 \\ 0 & (S_*'S)^{-1} \end{bmatrix} \begin{bmatrix} Q' \\ S_*' \end{bmatrix} \\ &= Q(Q'Q)^{-1}Q'+S_*(S_*'S_*)^{-1}S_*'. \end{aligned} \tag{27}$$

The result follows from the fact that $Z(Z'Z)^{-1}Z'=Z(Z^{-1}Z'^{-1})Z'=I$. It follows from (26) that the vector the ordinates of the polynomial regression is also given by

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

The use of polynomial regression in a preliminary detrending of the data is an essential part of a strategy for determining an appropriate representation of the underlying trajectory of an econometric data sequence. Once the trend has been eliminated from the data, one can proceed to assess their spectral structure by examining the periodogram of the residual sequence. Often the periodogram will reveal the existence of a cut-off frequency that bounds a low frequency trend/cycle component and separates it from the remaining elements of the spectrum.

An example is given in Figures 2 and 3. Figure 2 represents the logarithms of the quarterly data on aggregate consumption in the United Kingdom for the years 1955 to 1994. Through these data, a linear trend has been interpolated by least-squares regression. This line establishes a benchmark of constant

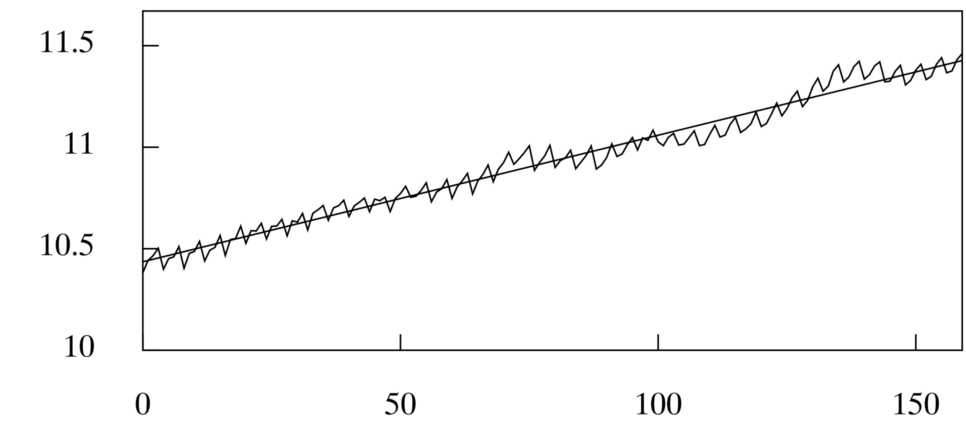

Figure 2. The quarterly series of the logarithms of consumption in the U.K., for the years 1955 to 1994, together with a linear trend interpolated by least-squares regression.

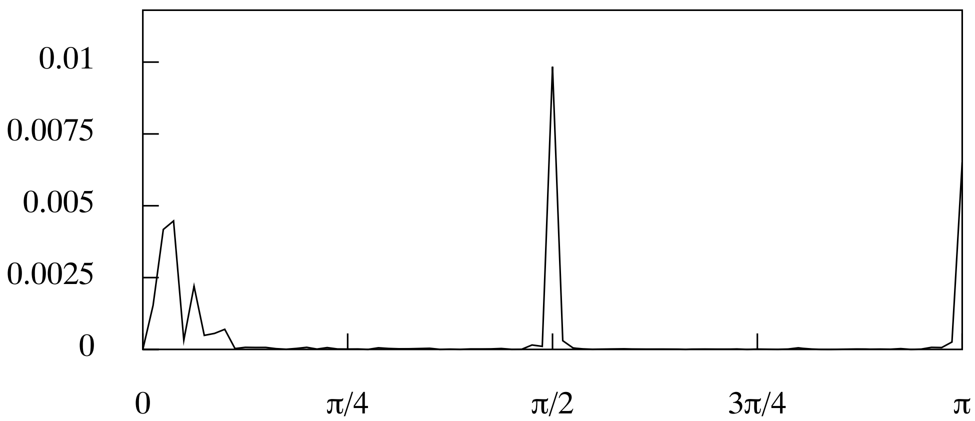

Figure 3 . The periodogram of the residual sequence obtained from the linear detrending of the logarithmic consumption data.

exponential growth, against which the fluctuations of consumption can be measured. The periodogram of the residual sequence in plotted in Figure 3. This shows that the low-frequency structure is bounded by a frequency value of $\pi/8$. This value can used in specifying the appropriate filter for extracting the low-frequency trajectory of the data.

For different data, a more flexible trend function might be appropriate. This could be provided by a polynomial of a higher degree. However, when a distinct cut-off frequency is revealed by the periodogram, it will be evident regardless of the degree of the polynomial.

The effect of increasing the degree of the trend polynomial will be to create a more flexible function. This will enable the trend to absorb some of the low-frequency fluctuations of the data, thereby diminishing their power within the residual sequence. The effect is to reduce the prominence, within the periodogram, of the low-frequency spectral structure.

The means of reducing data to stationarity that has been adopted traditionally by econometricians has been to take differences of the data. If the twofold difference operator for finite samples is represented by the matrix Q' and if y is the vector of the data, then the differenced data will be represented by $g = Q'y$. Reference to equation (28) will show that this contains the same information as does the vector $e = Q(Q'Q)^{-1}Q'y$ of the residuals from a linear detrending.

However, as Figure 1 implies, the effect of the difference operator is to attenuate the low-frequency fluctuations so severely as to render their spectral structure all but invisible. This problem does not affect the corresponding sequence of regression residuals.

Wiener–Kolmogorov filtering

The modern theory of statistical signal extraction was formulated independently by Wiener (1941) and Kolmogorov (1941), who arrived at the same results in different ways. Whereas Kolmogorov took a time-domain approach to the problem, Wiener worked primarily in the frequency domain. However, the unification of the two approaches was soon achieved.

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

$$y(t) = \xi(t) + \eta(t), \tag{29}$$

which is afflicted by the noise $\eta(t)$. According to the classical assumptions, which we shall later amend in order to accommodate short non-stationary sequences, the signal and the noise are generated by zero-mean stationary stochastic processes that are mutually independent. Also, the assumption is made that the data constitute a doubly-infinite sequence. It follows that the autocovariance generating function of the data is the sum of the autocovariance generating functions of its two components. Thus

$$\gamma^{yy}(z) = \gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z) \quad \text{and} \quad \gamma^{\xi\xi}(z) = \gamma^{y\xi}(z). \tag{30}$$

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

$$\gamma^{yy}(z) = \phi(z^{-1})\phi(z), \quad \gamma^{\xi\xi}(z) = \theta(z^{-1})\theta(z), \quad \gamma^{\eta\eta}(z) = \theta_\eta(z^{-1})\theta_\eta(z). \tag{31}$$

The estimate x_t of the signal element ξ_t is a linear combination of the elements of the data sequence:

$$x_t = \sum_j \beta_j y_{t-j}. \tag{32}$$

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, the following condition applies for all k:

$$\begin{aligned} 0 &= E\Big\{y_{t-k}(\xi_t - x_t)\Big\} \\ &= E(y_{t-k}\xi_t) - \sum_j \beta_j E(y_{t-k}y_{t-j}) \\ &= \gamma_k^{y\xi} - \sum_j \beta_j \gamma_{k-j}^{yy}. \end{aligned} \tag{33}$$

The equation may be expressed, in terms of the z-transforms, as

$$\gamma^{y\xi}(z) = \beta(z)\gamma^{yy}(z), \tag{34}$$

It follows that

$$\begin{aligned} \beta(z) &= \frac{\gamma^{y\xi}(z)}{\gamma^{yy}(z)} \\ &= \frac{\gamma^{\xi\xi}(z)}{\gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z)} = \frac{\theta(z^{-1})\theta(z)}{\rho(z^{-1})\rho(z)}. \end{aligned} \tag{35}$$

Now, by setting $z = \exp\{\mathrm{i}\omega\}$, one can derive the frequency-response function of the filter that is used in estimating the signal $\xi(t)$. The effect of the filter is to multiply each of the frequency elements of $y(t)$ by the fraction of its variance that is attributable to the signal. The same principle applies to the estimation of the residual component. This is obtained using the complementary filter

$$\beta^c(z) = 1 - \beta(z) = \frac{\gamma^{\eta\eta}(z)}{\gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z)}. \tag{36}$$

The estimated signal component may be obtained by filtering the data in two passes according to the following equations:

$$\phi(z)q(z) = \theta(z)y(z), \qquad \phi(z^{-1})x(z^{-1}) = \theta(z^{-1})q(z^{-1}). \tag{37}$$

The first equation relates to a process that runs forwards in time to generate the elements of an intermediate sequence, represented by the coefficients of

$q(z)$. The second equation represents a process that runs backwards to deliver the estimates of the signal, represented by the coefficients of $x(z)$.

The Wiener–Kolmogorov methodology can be applied to non stationary data with minor adaptations. A model of the processes underlying the data can be adopted that has the form of

$$\begin{aligned}\nabla^d(z)y(z) &= \nabla^d(z)\{\xi(z)+\eta(z)\} = \delta(z)+\kappa(z)\\ &= (1+z)^n\zeta(z)+(1-z)^m\varepsilon(z),\end{aligned} \tag{38}$$

where $\zeta(z)$ and $\varepsilon(z)$ are the z-transforms of two independent white-noise sequences $\zeta(t)$ and $\varepsilon(t)$. The condition $m \geq d$ is necessary to ensure the stationarity of $\eta(t)$, which is obtained from $\varepsilon(t)$ by differencing $m-d$ times. Then, the filter that is applied to $y(t)$ to estimate $\xi(t)$, which is the d-fold integral of $\delta(t)$, takes the form of

$$\beta(z) = \frac{\sigma_\zeta^2(1+z^{-1})^n(1+z)^n}{\sigma_\zeta^2(1+z^{-1})^n(1+z)^n+\sigma_\varepsilon^2(1-z^{-1})^m(1-z)^m}, \tag{39}$$

regardless of the degree d of differencing that would be necessary to reduce $y(t)$ to stationarity.

Two special cases are of interest. By setting $d=m=2$ and $n=0$ in (39), a model is obtained of a second-order random walk $\xi(t)$ affected by white-noise errors of observation $\eta(t)=\varepsilon(t)$. The resulting lowpass W–K filter, in the form of

$$\beta(z) = \frac{1}{1+\lambda(1-z^{-1})^2(1-z)^2} \quad \text{with} \quad \lambda = \frac{\sigma_\eta^2}{\sigma_\delta^2}, \tag{40}$$

is the Hodrick–Prescott (H–P) filter. The complementary highpass filter, which generates the residue, is

$$\beta^c(z) = \frac{(1-z^{-1})^2(1-z)^2}{\lambda^{-1}+(1-z^{-1})^2(1-z)^2}. \tag{41}$$

Here, λ, which is described as the smoothing parameter, is the single adjustable parameter of the filter.

By setting $m=n$, a filter for estimating $\xi(t)$ is obtained that takes the form of

$$\begin{aligned}\beta(z) &= \frac{\sigma_\zeta^2(1+z^{-1})^n(1+z)^n}{\sigma_\zeta^2(1+z^{-1})^n(1+z)^n+\sigma_\varepsilon^2(1-z^{-1})^n(1-z)^n}\\ &= \frac{1}{1+\lambda\left(\mathrm{i}\dfrac{1-z}{1+z}\right)^{2n}} \quad \text{with} \quad \lambda = \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}.\end{aligned} \tag{42}$$

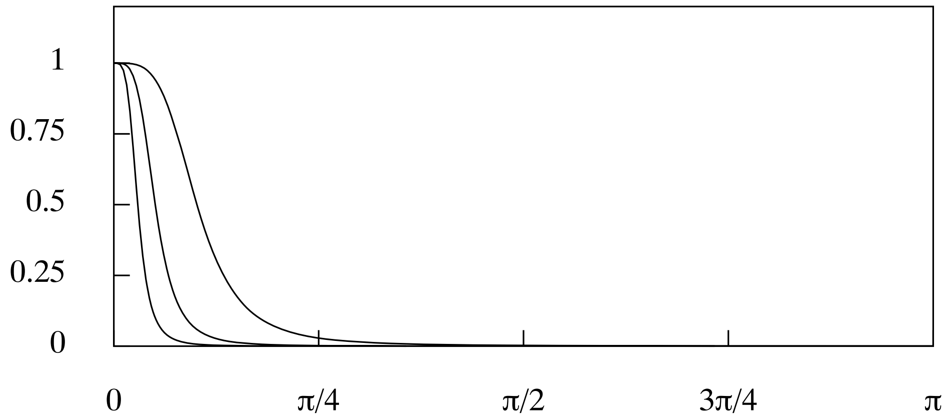

Figure 4. The gain of the Hodrick–Prescott lowpass filter with a smoothing parameter set to 100, 1600 and 14400.

This is the formula for the Butterworth lowpass digital filter. The filter has two adjustable parameters, and, therefore, it is a more flexible device than the H–P filter. First, there is the parameter λ. This can be expressed as

$$\lambda = \{1/\tan(\omega_d)\}^{2n}, \tag{43}$$

where ω_d is the nominal cut-off point of the filter, which is the mid point in the transition of the filter's frequency response from its pass band to its stop band. The second of the adjustable parameters is n, which denotes the order of the filter. As n increases, the transition between the pass band and the stop band becomes more abrupt.

These filters can be applied to the nonstationary data sequence $y(t)$ in the manner indicated by equation (37), provided that the appropriate initial conditions are supplied with which to start the recursions. However, by concentrating on the estimation of the residual sequence $\eta(t)$, which corresponds to a stationary process, it is possible to avoid the need for nonzero initial conditions. Then, the estimate of $\eta(t)$ can be subtracted from $y(t)$ to obtain the estimate of $\xi(t)$.

The H–P filter has been used as a lowpass smoothing filter in numerous macroeconomic investigations, where it has been customary to set the smoothing parameter to certain conventional values. Thus, for example, the econometric computer package *Eviews 4.0* (2000) imposes the following default values:

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

Figure 4 shows the square gain of the filter corresponding to these values. The innermost curve corresponds to $\lambda = 14,400$ and the outermost curve to $\lambda = 100$.

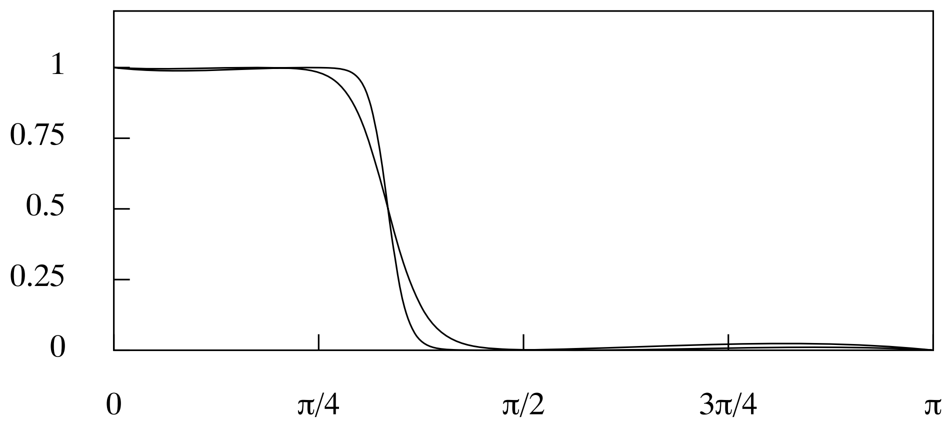

Figure 6. The squared gain of the lowpass Butterworth filters of orders $n = 6$ and $n = 12$ with a nominal cut-off point of $2\pi/3$ radians.

Whereas they have become conventional, these values are arbitrary. The filter should be adapted to the purpose of isolating the component of interest; and the appropriate filter parameters need to be determined in the light of the spectral structure of the component, such as has been revealed in Figure 3, in the case of the U.K. consumption data.

It will be observed that an H–P filter with $\lambda = 16,000$, which defines the middle curve in Figure 4, will not be effective in isolating the low-frequency component of the quarterly consumption data, which lies in the interval $[0, \pi]$. The curve will cut through the spectral structure; and the effect will be greatly to attenuate some of the elements of the component that should be preserved intact.

Lowering the value of λ in order to admit a wider range of frequencies will have the effect of creating a frequency response with a gradual transition from the pass band to the stop band. This will be equally inappropriate to the purpose of isolating a component within a well- defined frequency band. For that purpose, a different filter is required.

A filter that may be appropriate to the purpose of isolating the low-frequency fluctuations in consumption is the Butterworth filter. The squared gain of the latter is illustrated in Figure 5. In this case, there is a well-defined nominal cut-off frequency, which is a the mid point of the transition from the pass band to the stop band. The transition becomes more rapid as the filter order n increases. If a perfectly sharp transition is required, then the frequency-domain filter that will be presented later should be employed.

The Hodrick–Prescott filter has many antecedents. Its invention cannot reasonably be attributed to Hodrick and Prescott (1980, 1997), who cited Whittaker (1923) as one of their sources. Leser (1961) also provided a complete derivation of the filter at an earlier date. The Butterworth filter is a commonplace of electrical engineering.

4. The Finite-Sample Realisations of the W–K Filters

To derive the finite-sample version of a Wiener–Kolmogorov filter, we may consider a data vector $y = [y_0, y_1, \ldots, y_{t-1},]'$ that has a signal component ξ and a noise component η:

$$y = \xi + \eta. \tag{44}$$

The two components are assumed to be independently normally distributed with zero means and with positive-definite dispersion matrices. Then,

$$\begin{aligned} E(\xi) &= 0, \qquad D(\xi) = \Omega_\xi, \\ E(\eta) &= 0, \qquad D(\eta) = \Omega_\eta, \\ &\text{and} \quad C(\xi, \eta) = 0. \end{aligned} \tag{45}$$

The dispersion matrices Ω_ξ and Ω_η may be obtained, from the autocovariance generating functions $\gamma_\xi(z)$ and $\gamma_\eta(z)$, respectively, by replacing z by the matrix argument L_T, which is the finite sample version of the lag operator. Negative powers of z are replaced by powers of the forwards shift operator $F_T = L_T^{-1}$. A consequence of the independence of ξ and η is that $D(y) = \Omega_\xi + \Omega_\eta$.

The optimal predictors of the signal and the noise components are the following conditional expectations:

$$\begin{aligned} E(\xi|y) &= E(\xi) + C(\xi, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y = Z_\xi y = x, \end{aligned} \tag{46}$$

$$\begin{aligned} E(\eta|y) &= E(\eta) + C(\eta, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y = Z_\eta y = h, \end{aligned} \tag{47}$$

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

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

$$\begin{aligned} D(\xi|y) &= D(\xi) - C(\xi, y)D^{-1}(y)C(y, \xi) \\ &= \Omega_\xi - \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\xi, \end{aligned} \tag{48}$$

$$\begin{aligned} D(\eta|y) &= D(\eta) - C(\eta, y)D^{-1}(y)C(y, \eta), \\ &= \Omega_\eta - \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\eta. \end{aligned} \tag{49}$$

The estimates of ξ and η, which have been denoted by x and h respectively, can also be derived according to the following criterion:

$$\text{(50)} \qquad \text{Minimise} \quad S(\xi,\eta) = \xi'\Omega_\xi^{-1}\xi + \eta'\Omega_\eta^{-1}\eta \quad \text{subject to} \quad \xi + \eta = y.$$

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

Substituting for $\eta = y - \xi$ gives the concentrated criterion function $S(\xi) = \xi'\Omega_\xi^{-1}\xi + (y-\xi)'\Omega^{-1}(y-\xi)$. Differentiating this function in respect of ξ and setting the result to zero gives a condition for a minimum, which specifies the estimate x. This is $\Omega_\eta^{-1}(y-x) = \Omega_\xi^{-1}x$, which, on pre multiplication by Ω_η, can be written as $y = x - \Omega_\eta\Omega_\xi^{-1}x = (\Omega_\xi + \Omega_\eta)\Omega_\xi^{-1}x$. Therefore, the solution for x is

$$\text{(51)} \qquad x = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y.$$

Moreover, since the roles of ξ and η are interchangeable in this exercise, and, since $h + x = y$, there are also

$$\text{(52)} \qquad h = \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y \qquad \text{and} \qquad x = y - \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y.$$

The filter matrices $B_\xi = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}$ and $B_\eta = \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}$ of (51) and (52) are the matrix analogues of the z-transforms displayed in equations (35) and (36).

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

$$\text{(53)} \qquad (\Omega_\xi + \Omega_\eta)b = y$$

for the value of b. Thereafter, one can generate

$$\text{(54)} \qquad x = \Omega_\xi b \qquad \text{and} \qquad h = \Omega_\eta b.$$

If Ω_ξ and Ω_η correspond to the narrow-band dispersion matrices of moving-average processes, then the solution to equation (53) may be found via a Cholesky factorisation that sets $\Omega_\xi + \Omega_\eta = GG'$, where G is a lower-triangular matrix with a limited number of nonzero bands. The system $GG'b = y$ may be cast in the form of $Gp = y$ and solved for p. Then, $G'b = p$ can be solved for b.

Filters for Short Trended Sequences

One way of eliminating the trend is to take differences of the data. Usually, twofold differencing is appropriate. The matrix analogue of the second-order backwards difference operator in the case of $T = 5$ is given by

$$\nabla_5^2 = \begin{bmatrix} Q'_* \\ Q' \end{bmatrix} = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ \hline 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{array}\right]. \tag{55}$$

The first two rows, which do not produce true differences, are liable to be discarded. In general, the p-fold differences of a data vector of T elements will be obtained by pre multiplying it by a matrix Q' of order $(T-p) \times T$. Applying Q' to the equation $y = \xi + \eta$, representing the trended data, gives

$$\begin{aligned} Q'y &= Q'\xi + Q'\eta \\ &= \delta + \kappa = g. \end{aligned} \tag{56}$$

The vectors of expectations and the dispersion matrices of the differenced vectors are

$$\begin{aligned} E(\delta) &= 0, \qquad D(\delta) = \Omega_\delta = Q'D(\xi)Q, \\ E(\kappa) &= 0, \qquad D(\kappa) = \Omega_\kappa = Q'D(\eta)Q. \end{aligned} \tag{57}$$

The difficulty of estimating the trended vector $\xi = y - \eta$ directly is that some starting values or initial conditions are required in order to define the value at time $t = 0$. However, since η is from a stationary mean-zero process, it requires only zero-valued initial conditions. Therefore, the starting-value problem can be circumvented by concentrating on the estimation of η. The conditional expectation of η given the differenced data $g = Q'y$ is provided by the formula

$$\begin{aligned} h = E(\eta|g) &= E(\eta) + C(\eta, g)D^{-1}(g)\{g - E(g)\} \\ &= C(\eta, g)D^{-1}(g)g, \end{aligned} \tag{58}$$

where the second equality follows in view of the zero-valued expectations. Within this expression, there are

$$D(g) = \Omega_\delta + Q'\Omega_\eta Q \quad \text{and} \quad C(\eta, g) = \Omega_\eta Q. \tag{59}$$

Putting these details into (57) gives the following estimate of η:

$$h = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{60}$$

Putting this into the equation

$$x = E(\xi|g) = y - E(\eta|g) = y - h \tag{61}$$

gives

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

As in the case of the extraction of a signal from a stationary process, the estimate of the trended vector ξ can also be derived according to a least-squares criterion. The criterion is

$$\text{Minimise} \quad (y-\xi)'\Omega_\eta^{-1}(y-\xi) + \xi'Q\Omega_\delta^{-1}Q'\xi. \tag{63}$$

The first term in this expression penalises the departures of the resulting curve from the data, whereas the second term imposes a penalty for a lack of smoothness. Differentiating the function with respect to ξ and setting the result to zero gives

$$\Omega_\eta^{-1}(y-x) = -Q\Omega_\delta^{-1}Q'x = Q\Omega_\delta^{-1}d, \tag{64}$$

where x stands for the estimated value of ξ and $d = Q'x$. Premultiplying by $Q'\Omega_\eta$ gives

$$Q'(y-x) = Q'y - d = Q'\Omega_\eta Q\Omega_\delta^{-1}d, \tag{66}$$

whence

$$\begin{aligned} Q'y &= d + Q'\Omega_\eta Q\Omega_\delta^{-1}d \\ &= (\Omega_\delta + Q'\Omega_\eta Q)\Omega_\delta^{-1}d, \end{aligned} \tag{66}$$

which gives

$$\Omega_\delta^{-1}d = (\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{67}$$

Putting this into

$$x = y - \Omega_\eta Q\Omega_\delta^{-1}d, \tag{68}$$

which comes from premultiplying (63) by Ω_η, gives

$$x = y - \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{69}$$

One may observe that

$$\Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'e, \tag{70}$$

where $e = Q(Q'Q)^{-1}Q'y$ is the vector of residuals obtained by interpolating a straight line through the data by a least-squares regression. That is to say, it makes no difference to the estimate of the component that is complementary to the trend whether the filter is applied to the data vector y or the residual vector e. If the trend-estimation filter is applied to e instead of to y, then the resulting vector can be added to the ordinates of the interpolated line to create the estimate of the trend.

The specific cases that have been considered in the context of the classical form of the of Wiener–Kolmogorov filter can now be adapted to the circumstances of short trended sequences. The first there is the Leser filter. This is derived by setting

$$D(\eta) = \Sigma = \sigma_\eta^2 I, \quad D(\delta) = \sigma_\delta^2 I \quad \text{and} \quad \lambda = \frac{\sigma_\eta^2}{\sigma_\delta^2} \tag{71}$$

within (68) to give

$$x = y - Q(\lambda^{-1}I + Q'Q)^{-1}Q'y \tag{72}$$

Here, λ is the so-called smoothing parameter. It will be observed that, as $\lambda \to \infty$, the vector x tends to that of a linear function interpolated into the data by least-squares regression, which is represented by equation (28). The matrix expression $B = I - Q(\lambda^{-1}I + Q'Q)^{-1}Q'$ for the filter can be compared to the polynomial expression $\beta^c(z) = 1 - \beta(z)$ of the classical formulation, which entails the z-transform from (41).

The Butterworth filter that is appropriate to short trended sequences can be represented by the equation

$$x = y - \lambda\Sigma Q(M + \lambda Q'\Sigma Q)^{-1}Q'y. \tag{73}$$

Here, the matrices

$$\Sigma = \{2I_T - (L_T + L_T')\}^{n-2} \quad \text{and} \quad M = \{2I_T + (L_T + L_T')\}^n \tag{74}$$

are obtained from the RHS of the equations $\{(1-z)(1-z^{-1})\}^{n-2} = \{2 - (z + z^{-1})\}^{n-2}$ and $\{(1+z)(1+z^{-1})\}^n = \{2 + (z + z^{-1})\}^n$, respectively, by replacing z by L_T and z^{-1} by L_T'. Observe that the equalities no longer hold after the replacements. However, it can be verified that

$$Q'\Sigma Q = \{2I_T - (L_T + L_T')\}^n. \tag{75}$$

Filtering in the Frequency Domain

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

$$\begin{aligned} \Omega^{\circ}_{\xi} &= \bar{U}\gamma_{\xi}(D)U, \quad \Omega^{\circ}_{\eta} = \bar{U}\gamma_{\eta}(D)U \quad \text{and} \\ \Omega^{\circ} &= \Omega^{\circ}_{\xi} + \Omega^{\circ}_{\eta} = \bar{U}\{\gamma_{\xi}(D) + \gamma_{\eta}(D)\}U, \end{aligned} \tag{76}$$

wherein the diagonal matrices $\gamma_{\xi}(D)$ and $\gamma_{\eta}(D)$ contain the ordinates of the spectral density functions of the component processes.

Here, $U = T^{-1/2}[W^{jt}]$, wherein $t, j = 0, \ldots, T-1$, is the matrix of the Fourier transform, of which the generic element in the jth row and tth column is $W^{jt} = \exp(-\mathrm{i}2\pi tj/T)$, and $\bar{U}$ is its conjugate transpose. Also, $D = \mathrm{diag}\{1, W, W^2, \ldots, W^{T-1}\}$, which replaces z within each of the autocovariance generating functions, 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.

By replacing the dispersion matrices within (46) and (47) by their circulant counterparts, we derive the following formulae:

$$x = \bar{U}\gamma_{\xi}(D)\{\gamma_{\xi}(D) + \gamma_{\eta}(D)\}^{-1}Uy = P_{\xi}y, \tag{77}$$

$$h = \bar{U}\gamma_{\eta}(D)\{\gamma_{\xi}(D) + \gamma_{\eta}(D)\}^{-1}Uy = P_{\eta}y. \tag{78}$$

Similar replacements within the formulae (48) and (49) 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 $\bar{U}$.

The filters described above are appropriate only to stationary processes. However, they can be adapted in several alternative ways to cater to nonstationary processes. One way is to reduce the data may to stationarity by twofold differencing before filtering it. After filtering, the data may be reinflated by a process of summation.

As before, let the original data be denoted by $y = \xi + \eta$ and let the differenced data be $g = Q'y = \delta + \kappa$. If the estimates of $\delta = Q'\xi$ and $\kappa = Q'\eta$ are denoted by d and k respectively, then the estimates of ξ and η will be

$$x = S_*d_* + Sd \quad \text{where} \quad d_* = (S_*'S_*)^{-1}S_*'(y - Sd) \tag{79}$$

and

$$h = S_* k_* + Sk \quad \text{where} \quad k_* = -(S_*' S_*)^{-1} S_*' Sk. \tag{80}$$

Here, d_* an k_* are the initial conditions that are obtained via the minimisation of the function

$$\begin{aligned}(y - x)'(y - x) &= (y - S_* d_* - Sd)'(y - S_* d_* - Sd) \\ &= (S_* k_* + Sk)'(S_* k_* + Sk) = h'h.\end{aligned} \tag{81}$$

The minimisation ensures that the estimated trend x adheres as closely as possible to the data y.

In the case where the data is differenced twice, there is

$$S_*' = \begin{bmatrix} 1 & 2 & \ldots & T-1 & T \\ 0 & 1 & \ldots & T-2 & T-1 \end{bmatrix} \tag{82}$$

The elements of the matrix $S_*' S_*$ can be found via the formulae

$$\begin{aligned}\sum_{t=1}^{T} t^2 &= \frac{1}{6} T(T+1)(2T+1) \quad \text{and} \\ \sum_{t=1}^{T} t(t-1) &= \frac{1}{6} T(T+1)(2T+1) - \frac{1}{2} T(T+1).\end{aligned} \tag{83}$$

(A compendium of such results has been provided by Jolly 1961, and proofs of the present results were given by Hall and Knight 1899.)

Example. Before applying a frequency-domain filter, it is necessary to ensure that the data are free of trend. If a trend is detected, then it may be removed from the data by subtracting an interpolated polynomial trend function. A test for the presence of a trend is required that differs from the tests that are used to detect the presence of unit roots in the processes generating the data. This is provided by the significance test associated with the ordinary-least squares estimate of a linear trend.

There is a simple means of calculating the adjusted sum of squares of the temporal index $t = 0, 1, \ldots, T-1$, which is entailed in the calculation of the slope coefficient

$$b = \frac{\sum y_t^2 - (\sum y_t)^2 / T}{\sum t^2 - (\sum t)^2 / T}. \tag{84}$$

The formulae

$$\sum_{t=0}^{T-1} t^2 = \frac{1}{6}(T-1)T(2T-1) \quad \text{and} \quad \sum_{t=0}^{T-1} t = \frac{T(T-1)}{2} \tag{85}$$

are combined to provide a convenient means of calculating the denominator of the formula of (82):

$$\sum_{t=0}^{T-1} t^2 - \frac{(\sum_{t=0}^{T-1} t)^2}{T} = \frac{(T-1)T(T+1)}{12}. \tag{86}$$

Another means of calculating the low-frequency trajectory of the data via the frequency domain mimics the method of equation (69) by concentrating of the estimation the high-frequency component. This can be subtracted from the data to create an estimate of the complementary low-frequency trend component. However, whereas, in the case of equation (69), the differencing of the data and the re-inflation of the estimated high-frequency component are deemed to take place in the time domain now the re-inflation occurs in the frequency domain before the resulting vector of Fourier coefficients is transformed to the time domain.

The reduction of trended data sequence to stationary continues to be effected by the matrix Q but, in this case, the matrix can be seen in the context of a centralised difference operator This is

$$\begin{aligned} N(z) &= z^{-1} - 2 + z = z^{-1}(1-z)^2 \\ &= z^{-1}\nabla^2(z). \end{aligned} \tag{87}$$

The matrix version of the operator is obtained by setting $z = L_T$ and $z^{-1} = L_T'$, which gives

$$N(L_T) = N_T = L_T - 2I_T + L_T'. \tag{88}$$

The first and the final rows of this matrix do not deliver true differences. Therefore, they are liable to be deleted, with the effect that the two end points are lost from the twice-differenced data. Deleting the rows $e_0'N_T$ and $e_{T-1}'N_T$ from N_T gives the matrix Q', which can also be obtained by from $\nabla_T^2 = (I_T - L_T)^2$ by deleting the matrix Q_*', which comprises the first two rows $e_0'\nabla_T^2$ and $e_1'\nabla_T^2$. In the case of $T = 5$ there is

$$N_5 = \begin{bmatrix} Q_{-1}' \\ Q' \\ Q_{+1} \end{bmatrix} = \left[\begin{array}{ccccc} -2 & 1 & 0 & 0 & 0 \\ \hline 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ \hline 0 & 0 & 0 & 1 & -2 \end{array}\right]. \tag{89}$$

On deleting the first and last elements of the vector N_Ty, which are $Q_{-1}'y = e_1'\nabla_T^2 y$ and $Q_{+1}y$, respectively, we get $Q'y = [q_1, \ldots, q_{T-2}]'$.

The loss of the two elements from either end of the (centrally) twice-differenced data can be overcome by supplementing the original data vector y with two extrapolated end points y_{-1} and y_T. Alternatively, the differenced data may be supplemented by attributing appropriate values to q_0 and q_{T-1}. These could be zeros or some combination of the adjacent values. In either case, we will obtain a vector of order T denoted by $q = [q_0, q_1, \ldots q_{T-1}]'$.

In describing the method for implementing a highpass filter, let Λ be the matrix which selects the appropriate ordinates of the Fourier transform $\gamma = Uq$ of the twice differenced data. These ordinates must be reinflated to compensate for the differencing operation, which has the frequency response

(90) $$f(\omega) = 2 - 2\cos(\omega).$$

The response of the anti-differencing operation is $1/f(\omega)$; and γ is reinflated by pre-multiplying by the diagonal matrix

(91) $$V = \mathrm{diag}\{v_0, v_1, \ldots, v_{T-1}\},$$

comprising the values $v_j = 1/f(\omega_j); j = 0, \ldots, T-1$, where $\omega_j = 2\pi j/T$.

Let $H = V\Lambda$ be the matrix that is is applied to $\gamma = Uq$ to generate the Fourier ordinates of the filtered vector. The resulting vector is transformed to the time domain to give

(92) $$h = \bar{U}H\gamma = \bar{U}HUq.$$

It will be see that $f(\omega)$ is zero-valued when $\omega = 0$ and that $1/f(\omega)$ is unbounded in the neighbourhood of $\omega = 0$. Therefore, a frequency-domain reinflation is available only when there are no nonzero Fourier ordinates in this neighbourhood. That is to say, it can work only in conjunction with highpass or bandpass filtering. However, it is straightforward to construct a lowpass filter that complements the highpass filter. The low-frequency trend component that is complementary to h is

(93) $$x = y - h = y - \bar{U}HUq.$$