Statistical Signal Extraction and Filtering

by D.S.G. Pollock
University of Leicester
Email: stephen_pollock@sigmapi.u-net.com

Linear Time Invariant Filters

Whenever we form a linear combination of successive elements of a discrete-time signal $x(t) = \{x_t; t = 0, \pm 1, \pm 2, \ldots\}$, we are performing an operation that is described as linear filtering. In the case of a linear time-invariant filter, such an operation can be represented by the equation

$$y(t) = \sum_j \psi_j x(t - j). \tag{1}$$

To assist in the algebraic manipulation of such equations, we may convert the infinite sequences $x(t)$ and $y(t)$ and the sequence of filter coefficients $\{\psi_j\}$ into power series or polynomials. By associating z^t to each element y_t and by summing the sequence, we get

$$\sum_t y_t z^t = \sum_t \Big\{ \sum_j \psi_j x_{t-j} \Big\} z^t \quad \text{or} \quad y(z) = \psi(z) x(z), \tag{2}$$

where

$$x(z) = \sum_t x_t z^t, \quad y(z) = \sum_t y_t z^t \quad \text{and} \quad \psi(z) = \sum_j \psi_j z^j. \tag{3}$$

The convolution operation of equation (1) becomes an operation of polynomial multiplication in equation (2). We are liable to describe the z-transform $\psi(z)$ of the filter coefficients as the transfer function of the filter.

For a treatise on the z-transform, see Jury (1964).

The Impulse Response

The sequence $\{\psi_j\}$ of the filter's coefficients constitutes its response, on the output side, to an input in the form of a unit impulse. If the sequence is finite, then $\psi(z)$ is described as a moving-average filter or as a finite impulse-response

(FIR) filter. When the filter produces an impulse response of an indefinite duration, it is called an infinite impulse-response (IIR) filter. The filter is said to be causal or backward-looking if none of its coefficients is associated with a negative power of z. In that case, the filter is available for real-time signal processing.

Causal Filters

A practical filter, which is constructed from a limited number of components of hardware or software, must be capable of being expressed in terms of a finite number of parameters. Therefore, linear IIR filters which are causal will invariably entail recursive equations of the form

$$\sum_{j=0}^{p} \phi_j y_{t-j} = \sum_{j=0}^{q} \theta_j x_{t-j}, \quad \text{with} \quad \phi_0 = 1, \tag{4}$$

of which the z-transform is

$$\phi(z)y(z) = \theta(z)x(z), \tag{5}$$

wherein $\phi(z) = \phi_0 + \phi_1 z + \cdots + \phi_p z^p$ and $\theta(z) = \theta_0 + \theta z + \cdots + \theta_q z^q$ are finite-degree polynomials. The leading coefficient of $\phi(z)$ may be set to unity without loss of generality; and thus the output sequence $y(t)$ in equation (4) becomes a function not only of past and present inputs but also of past outputs, which are described as feedback.

The recursive equation may be assimilated to the equation under (2) by writing it in rational form:

$$y(z) = \frac{\theta(z)}{\phi(z)} x(z) = \psi(z)x(z). \tag{6}$$

On the condition that the filter is stable, the expression $\psi(z)$ stands for the series expansion of the ratio of the polynomials.

The stability of a rational transfer function $\theta(z)/\phi(z)$ can be investigated via its partial-fraction decomposition, which gives rise to a sum of simpler transfer functions that can be analysed readily. If the degree of the numerator of $\theta(z)$ exceeds that of the denominator $\phi(z)$, then long division can be used to obtain a quotient polynomial and a remainder that is a proper rational function. The quotient polynomial will correspond to a stable transfer function; and the remainder will be the subject of the decomposition.

Assume that $\theta(z)/\phi(z)$ is a proper rational function in which the denominator is factorised as

$$\phi(z) = \prod_{j=1}^{r} (1 - z/\lambda_j)^{n_j}, \tag{7}$$

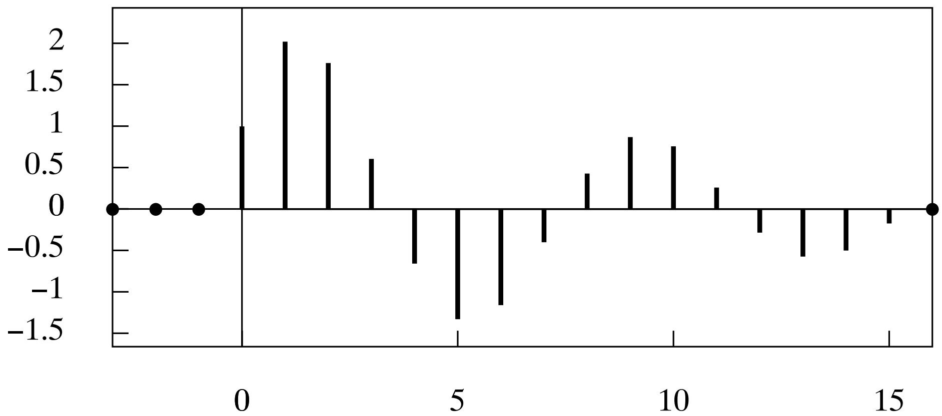

Figure 1. The impulse response of the transfer function $\theta(z)/\phi(z)$ with $\phi(z) = 1.0 - 1.2728z + 0.81z^2$ and $\theta(z)(z) = 1.0 + 0.0.75z$.

where n_j is the multiplicity of the root λ_j, and where $\sum_j n_j = p$ is the degree of the polynomial. Then, the so-called Heaviside partial-fraction decomposition is

$$\frac{\theta(z)}{\phi(z)} = \sum_{j=1}^{r}\sum_{k=1}^{n_j} \frac{c_{jk}}{(1 - z/\lambda_j)^k}; \tag{9}$$

and the task is to find the series expansions of the partial fractions. The stability of the transfer function depends upon the converge of the expansions. For this, the necessary and sufficient condition is that $|\lambda_j| > 1$ for all j, which is to to say that all of the roots of the denominator polynomial must lie outside the unit circle in the complex plane.

The expansions of a pair of partial fractions with conjugate complex roots will combine to produce a sinusoidal sequence. The expansion of a partial fraction containing a root of multiplicity n will be equivalent to the n-fold auto-convolution of the expansion a simple fraction containing the root.

It is helpful to represent the poles of the transfer function graphically by showing their locations within the complex plane together with the locations of the roots of the numerator polynomial, which are described as the zeros of the transfer function.

It is more convenient to represent the poles and zeros of $\theta(z^{-1})/\phi(z^{-1})$, which are the reciprocals of those of $\theta(z)/\phi(z)$. For a stable and invertible transfer function, these must lie within the unit circle. This recourse has been adopted for Figure 2, which shows the pole–zero diagram for the transfer function that gives rise to Figure 1.

The Series Expansion of a Rational Transfer Function

The method of finding the coefficients of the series expansion can be illus-

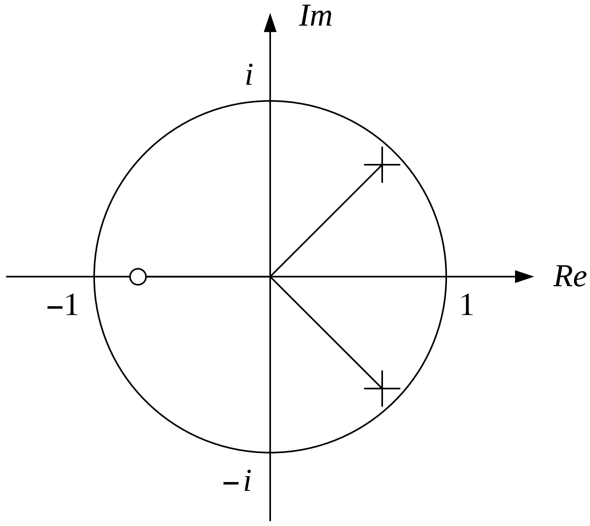

Figure 2. The pole–zero diagram corrresponding to the transfer function of Figure 1. The poles are conjugate complex numbers with arguments of $\pm\pi/4$ and with a modulus of 0.9. The single real-valued zero has the value of -0.75.

trated by the second-order case:

$$\text{(9)} \qquad \frac{\theta_0 + \theta_1 z}{\phi_0 + \phi_1 z + \phi_2 z^2} = \{\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots\}.$$

We rewrite this equation as

$$\text{(10)} \qquad \theta_0 + \theta_1 z = \{\phi_0 + \phi_1 z + \phi_2 z^2\}\{\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots\}.$$

The following table assists us in multiplying together the two polynomials:

(11)

	ψ_0	$\psi_1 z$	$\psi_2 z^2$	$\cdots$
ϕ_0	$\phi_0\psi_0$	$\phi_0\psi_1 z$	$\phi_0\psi_2 z^2$	$\cdots$
$\phi_1 z$	$\phi_1\psi_0 z$	$\phi_1\psi_1 z^2$	$\phi_1\psi_2 z^3$	$\cdots$
$\phi_2 z^2$	$\phi_2\psi_0 z^2$	$\phi_2\psi_1 z^3$	$\phi_2\psi_2 z^4$	$\cdots$

Performing the multiplication on the RHS of the equation, and by equating the coefficients of the same powers of z on the two sides, we find that

$$\text{(12)} \qquad \begin{aligned} \theta_0 &= \phi_0\psi_0, & \psi_0 &= \theta_0/\phi_0, \\ \theta_1 &= \phi_0\psi_1 + \phi_1\psi_0, & \psi_1 &= (\theta_1 - \phi_1\psi_0)/\phi_0, \\ 0 &= \phi_0\psi_2 + \phi_1\psi_1 + \phi_2\psi_0, & \psi_2 &= -(\phi_1\psi_1 + \phi_2\psi_0)/\phi_0, \\ &\vdots & &\vdots \\ 0 &= \phi_0\psi_n + \phi_1\psi_{n-1} + \phi_2\psi_{n-2}, & \psi_n &= -(\phi_1\psi_{n-1} + \phi_2\psi_{n-2})/\phi_0. \end{aligned}$$

Bi-directional (Non causal) Filters

A two-sided symmetric filter in the form of

$$\psi(z) = \theta(z^{-1})\theta(z) = \psi_0 + \psi_1(z^{-1} + z) + \cdots + \psi_m(z^{-m} + z^m) \tag{13}$$

is often employed in smoothing the data or in eliminating its seasonal components. The advantage of such a filter is the absence of a phase effect. That is to say, no delay is imposed on any of the components of the signal.

The so-called Cramér–Wold factorisation, which sets $\psi(z) = \theta(z^{-1})\theta(z)$, and which must be available for any properly-designed filter, provides a straightforward way of explaining the absence of a phase effect. The factorisation gives rise to two equations (i) $q(z) = \theta(z)y(z)$ and (ii) $x(z) = \theta(z^{-1})q(z)$. Thus, the transformation of (1) to be broken down into two operations:

$$\text{(i)} \quad q_t = \sum_j \theta_j y_{t-j} \qquad \text{and} \qquad \text{(ii)} \quad x_t = \sum_j \theta_j q_{t+j}. \tag{14}$$

The first operation, which runs in real time, imposes a time delay on every component of $x(t)$. The second operation, which works in reversed time, imposes an equivalent reverse-time delay on each component. The reverse-time delays, which are advances in other words, serve to eliminate the corresponding real-time delays.

If $\psi(z)$ corresponds to an FIR filter, then the processed sequence $x(t)$ may be generated via a single application of the two-sided filter $\psi(z)$ to the signal $y(t)$, or it may be generated in two operations via the successive applications of $\theta(z)$ to $y(z)$ and $\theta(z^{-1})$ to $q(z) = \theta(z)y(z)$. The question of which of these techniques has been used to generate $y(t)$ in a particular instance should be a matter of indifference.

The final species of linear filter that may be used in the processing of economic time series is a symmetric two-sided rational filter of the form

$$\psi(z) = \frac{\theta(z^{-1})\theta(z)}{\phi(z^{-1})\phi(z)}. \tag{15}$$

Such a filter must, of necessity, be applied in two separate passes running forwards and backwards in time and described, respectively, by the equations

$$\text{(i)} \quad \phi(z)q(z) = \theta(z)y(z) \qquad \text{and} \qquad \text{(ii)} \quad \phi(z^{-1})x(z) = \theta(z^{-1})q(z). \tag{16}$$

Such filters represent a most effective way of processing economic data in pursuance of a wide range of objectives.

The Response to a Sinusoidal Input

One must also consider the response of the transfer function to a simple sinusoidal signal. Any finite data sequence can be expressed as a sum of discretely

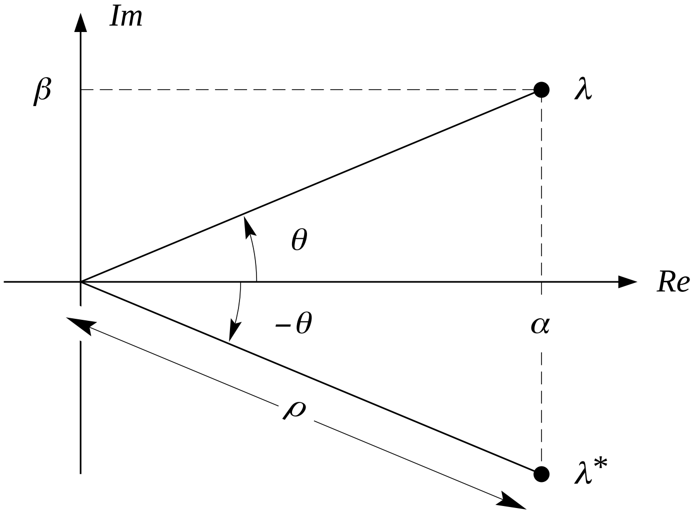

Figure 3. The Argand Diagram showing a complex number $\lambda = \alpha + i\beta$ and its conjugate $\lambda^* = \alpha - i\beta$.

sampled sine and cosine functions with frequencies that are integer multiples of a fundamental frequency that produces one cycle in the period spanned by the sequence. The finite sequence may be regarded as a single cycle within a infinite sequence, which is the periodic extension of the data.

Consider, therefore, the consequences of mapping the perpetual signal sequence $\{x_t = \cos(\omega t)\}$ through the transfer function with the coefficients $\{\psi_0, \psi_1, \ldots\}$. The output is

$$y(t) = \sum_j \psi_j \cos\big(\omega[t-j]\big). \tag{17}$$

By virtue of the trigonometrical identity $\cos(A-B) = \cos A \cos B + \sin A \sin B$, this becomes

$$\begin{aligned} y(t) &= \Big\{\sum_j \psi_j \cos(\omega j)\Big\}\cos(\omega t) + \Big\{\sum_j \psi_j \sin(\omega j)\Big\}\sin(\omega t) \\ &= \alpha\cos(\omega t) + \beta\sin(\omega t) = \rho\cos(\omega t - \theta), \end{aligned} \tag{18}$$

Observe that using the trigonometrical identity to expand the final expression of (16) gives $\alpha = \rho\cos(\theta)$ and $\beta = \rho\sin(\theta)$. Therefore,

$$\rho^2 = \alpha^2 + \beta^2 \qquad \text{and} \qquad \theta = \tan^{-1}\Big(\frac{\beta}{\alpha}\Big). \tag{19}$$

Also, if $\lambda = \alpha + \mathrm{i}\beta$ and $\lambda^* = \alpha - \mathrm{i}\beta$ are conjugate complex numbers, then ρ would be their modulus. This is illustrated in Figure 3.

It can be seen that the transfer function has a twofold effect upon the signal. First, there is a *gain effect,* whereby the amplitude of the sinusoid is

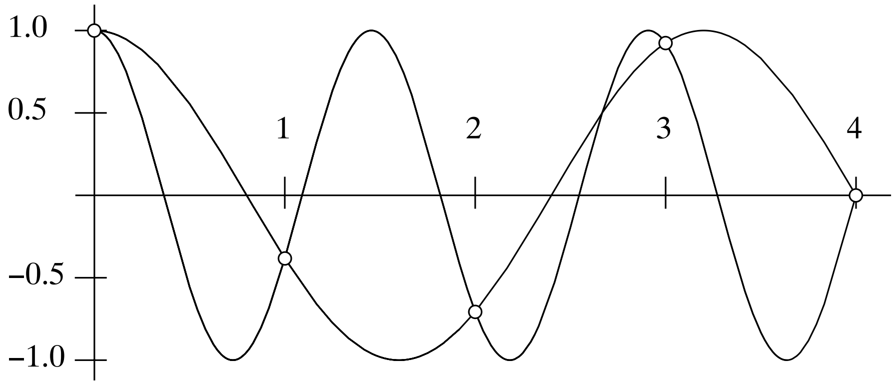

Figure 4. The values of the function $\cos\{(11/8)\pi t\}$ coincide with those of its alias $\cos\{(5/8)\pi t\}$ at the integer points $\{t = 0, \pm 1, \pm 2, \ldots\}$.

increased or diminished by the factor ρ. Then, there is a *phase effect,* whereby the peak of the sinusoid is displaced by a time delay of θ/ω periods. The frequency of the output is the same as the frequency of the input, which is a fundamental feature of all linear dynamic systems.

Observe that the response of the transfer function to a sinusoid of a particular frequency is akin to the response of a bell to a tuning fork. It gives very limited information regarding the characteristics of the system. To obtain full information, it is necessary to excite the system over a full range of frequencies.

Aliasing and the Shannon–Nyquist Sampling Theorem

In a discrete-time system, there is a problem of aliasing whereby signal frequencies (i.e. angular velocities) in excess of π radians per sampling interval are confounded with frequencies within the interval $[0, \pi]$. To understand this, consider a cosine wave of unit amplitude and zero phase with a frequency ω in the interval $\pi < \omega < 2\pi$ that is sampled at unit intervals. Let $\omega^* = 2\pi - \omega$. Then,

$$\begin{aligned} \cos(\omega t) &= \cos\big\{(2\pi - \omega^*)t\big\} \\ &= \cos(2\pi)\cos(\omega^* t) + \sin(2\pi)\sin(\omega^* t) \\ &= \cos(\omega^* t); \end{aligned} \tag{20}$$

which indicates that ω and ω^* are observationally indistinguishable. Here, $\omega^* \in [0, \pi]$ is described as the alias of $\omega > \pi$.

The maximum frequency in discrete data is π radians per sampling interval and, as the Shannon–Nyquist sampling theorem indicates, aliasing is avoided only if there are at least two observations in the time that it takes the signal component of highest frequency to complete a cycle. In that case, the discrete representation will contain all of the available information on the system.

The consequences of sampling at an insufficient rate are illustrated in Figure 4. Here, a rapidly alternating cosine function is mistaken for one of less than half the true frequency.

The sampling theorem is attributable to a several people, but it is most commonly attributed to Shannon (1949, 1989), albeit that Nyquist (1928) discovered the essential results at an earlier date.

The Frequency Response of a Linear Filter

The frequency response of a linear filter $\psi(z)$ is its response to the set of sinusoidal inputs of all frequencies ω that fall within the Nyquist interval $[0, \pi]$. This entails the squared gain of the filter, defined by

$$\rho^2(\omega) = \psi_\alpha^2(\omega) + \psi_\beta^2(\omega), \tag{21}$$

where

$$\psi_\alpha(\omega) = \sum_j \psi_j \cos(\omega j) \quad \text{and} \quad \psi_\beta(\omega) = \sum_j \psi_j \sin(\omega j), \tag{22}$$

and the phase displacement, defined by

$$\theta(\omega) = \text{Arg}\{\psi(\omega)\} = \tan^{-1}\{\psi_\beta(\omega)/\psi_\alpha(\omega)\}. \tag{23}$$

It is convenient to replace the trigonometrical functions of (20) by the complex exponential functions

$$e^{\mathrm{i}\omega j} = \frac{1}{2}\{\cos(\omega j) + \sin(\omega j)\} \quad \text{and} \quad e^{-\mathrm{i}\omega j} = \frac{1}{2}\{\cos(\omega j) - \sin(\omega j)\}, \tag{24}$$

which enable the trigonometrical functions to be expressed as

$$\cos(\omega t) = \frac{1}{2}\{e^{\mathrm{i}\omega j} + e^{-\mathrm{i}\omega j}\} \quad \text{and} \quad \sin(\omega j) = \frac{\mathrm{i}}{2}\{e^{-\mathrm{i}\omega j} - e^{\mathrm{i}\omega j}\}. \tag{25}$$

Setting $z = \exp\{-\mathrm{i}\omega j\}$ in $\psi(z)$ gives

$$\psi(e^{-\mathrm{i}\omega j}) = \psi_\alpha(\omega) - \mathrm{i}\psi_\beta(\omega), \tag{26}$$

which we shall write hereafter as $\psi(\omega) = \psi(e^{-\mathrm{i}\omega j})$.

The squared gain of the filter, previously denoted by $\rho^2(\omega)$, is the square of the complex modulus:

$$|\psi(\omega)|^2 = \psi_\alpha^2(\omega) + \psi_\beta^2(\omega), \tag{27}$$

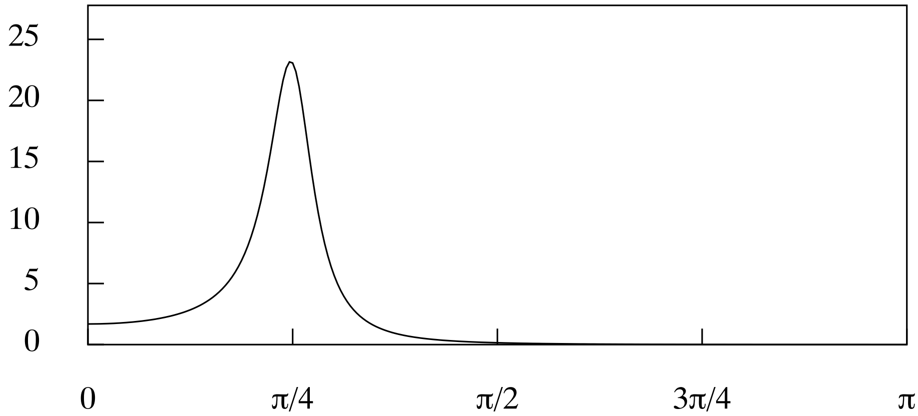

Figure 5. The spectral density function of the ARMA(2, 1) process $y(t) = 1.2728y(t-1) - 0.81y(t-2) + \varepsilon(t) + 0.0.75\varepsilon(t-1)$ with $V\{\varepsilon(t)\} = 1$.

which is obtained by setting $z = \exp\{-i\omega j\}$ in $\psi(z^{-1})\psi(z)$.

The Spectrum of a Stationary Stochastic Process

Consider a stationary stochastic process $y(t) = \{y_t; t = 0, \pm 1, \pm 2, \ldots\}$ defined on a doubly-infinite index set. The generic element of the process can be expressed as $y_t = \sum_j \psi_j \varepsilon_{t-j}$, where ε_t is an element of a sequence $\varepsilon(t)$ of independently and identically distributed random variables with $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = \sigma^2$ for all t.

The autocovariance generating function of the process is

$$\sigma^2\psi(z^{-1})\psi(z) = \gamma(z) = \{\gamma_0 + \gamma_1(z^{-1} + z) + \gamma_2(z^{-2} + z^2) + \cdots\}. \tag{28}$$

The following table assists us in forming the product $\gamma(z) = \sigma^2\psi(z^{-1})\psi(z)$:

$$\begin{array}{c|cccc} & \psi_0 & \psi_1 z & \psi_2 z^2 & \cdots \\ \hline \psi_0 & \psi_0^2 & \psi_0\psi_1 z & \psi_0\psi_2 z^2 & \cdots \\ \psi_1 z^{-1} & \psi_1\psi_0 z^{-1} & \psi_1^2 & \psi_1\psi_2 z & \cdots \\ \psi_2 z^{-2} & \psi_2\psi_0 z^{-2} & \psi_2\psi_1 z^{-1} & \psi_2^2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{array} \tag{29}$$

The autocovariances are obtained by summing along the NW–SE diagonals:

$$\begin{aligned} \gamma_0 &= \sigma^2\{\psi_0^2 + \psi_1^2 + \psi_2^2 + \psi_3^2 + \cdots\}, \\ \gamma_1 &= \sigma^2\{\psi_0\psi_1 + \psi_1\psi_2 + \psi_2\psi_3 + \cdots\}, \\ \gamma_2 &= \sigma^2\{\psi_0\psi_2 + \psi_1\psi_3 + \psi_2\psi_4 + \cdots\}, \\ &\vdots \end{aligned} \tag{30}$$

By setting $z = \exp\{-\mathrm{i}\omega j\}$ and dividing by 2π, we get the spectral density function of the process:

$$f(\omega) = \frac{1}{2\pi}\left\{\gamma_0 + 2\sum_{\tau=1}^{\infty}\gamma_\tau \cos(\omega\tau)\right\}. \tag{31}$$

This entails the cosine Fourier transform of the sequence of autocovariances.

The spectral density functions of an ARMA (2, 1) process, which incorporates the transfer function of Figures 1–3, is shown in Figure 5.

Wiener–Kolmogorov Filtering of Stationary Sequences

The classical theory of linear filtering was formulated independently by Norbert Weiner (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.

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

$$y(t) = \xi(t) + \eta(t), \tag{32}$$

which is afflicted by the noise $\eta(t)$. According to the classical assumptions, which we shall later amend in order to accommodate short nonstationary 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,

$$\gamma^{yy}(z) = \gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z) \quad \text{and} \quad \gamma^{\xi\xi}(z) = \gamma^{y\xi}(z). \tag{33}$$

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

$$\gamma^{yy}(z) = \phi(z^{-1})\phi(z), \quad \gamma^{\xi\xi}(z) = \theta(z^{-1})\theta(z), \quad \gamma^{\eta\eta}(z) = \theta_\eta(z^{-1})\theta_\eta(z). \tag{34}$$

The estimate x_t of the signal element ξ_t, generated by a linear time-invariant filter, is a linear combination of the elements of the data sequence:

$$x_t = \sum_j \psi_j y_{t-j}. \tag{35}$$

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:

$$\begin{aligned} 0 &= E\Big\{y_{t-k}(\xi_t - x_t)\Big\} \\ &= E(y_{t-k}\xi_t) - \sum_j \psi_j E(y_{t-k}y_{t-j}) \\ &= \gamma_k^{y\xi} - \sum_j \psi_j \gamma_{k-j}^{yy}. \end{aligned} \tag{36}$$

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

$$\gamma^{y\xi}(z) = \psi(z)\gamma^{yy}(z). \tag{37}$$

It follows that

$$\begin{aligned} \psi(z) &= \frac{\gamma^{y\xi}(z)}{\gamma^{yy}(z)} \\ &= \frac{\gamma^{\xi\xi}(z)}{\gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z)} = \frac{\theta(z^{-1})\theta(z)}{\phi(z^{-1})\phi(z)}. \end{aligned} \tag{38}$$

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

$$\psi^c(z) = 1 - \psi(z) = \frac{\gamma^{\eta\eta}(z)}{\gamma^{\xi\xi}(z) + \gamma^{\eta\eta}(z)}. \tag{39}$$

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

$$\phi(z)q(z) = \theta(z)y(z), \qquad \phi(z^{-1})x(z^{-1}) = \theta(z^{-1})q(z^{-1}). \tag{40}$$

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 Hodrick–Prescott (Leser) Filter and the Butterworth Filter

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

$$\begin{aligned} \nabla^d(z)y(z) &= \nabla^d(z)\{\xi(z) + \eta(z)\} = \delta(z) + \kappa(z) \\ &= (1+z)^n\zeta(z) + (1-z)^m\varepsilon(z), \end{aligned} \tag{41}$$

where $\zeta(z)$ and $\varepsilon(z)$ are the z-transforms of two independent white-noise sequences $\zeta(t)$ and $\varepsilon(t)$ and where $\nabla = 1 - z$ is the z-transform of the difference operator.

The model of $y(t) = \xi(t) + \eta(t)$ entails a pair of stochastic processes, which are defined over the doubly-infiinite sequence of integers and of which the z-transforms are

$$\xi(z) = \frac{(1+z)^n}{\nabla^d(z)}\zeta(z) \quad \text{and} \quad \eta(z) = \frac{(1-z)^m}{\nabla^d(z)}\varepsilon(z). \tag{42}$$

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.

It must be conceded that a nonstationary process such as $\xi(t)$ is a mathematical construct of doubtful reality, since its values will be unbounded, almost certainly. Nevertheless, to deal in these terms is to avoid the complexities of the finite-sample approach, which will be the subject of the next section.

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

$$\psi(z) = \frac{\sigma_\zeta^2(1+z^{-1})^n(1+z)^n}{\sigma_\zeta^2(1+z^{-1})^n(1+z)^n + \sigma_\varepsilon^2(1-z^{-1})^m(1-z)^m}, \tag{43}$$

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 (41), 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

$$\psi(z) = \frac{1}{1 + \lambda(1-z^{-1})^2(1-z)^2} \quad \text{with} \quad \lambda = \frac{\sigma_\eta^2}{\sigma_\delta^2}, \tag{44}$$

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

$$\psi^c(z) = \frac{(1-z^{-1})^2(1-z)^2}{\lambda^{-1} + (1-z^{-1})^2(1-z)^2}. \tag{45}$$

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

$$\begin{aligned}\psi(z) &= \frac{\sigma_\zeta^2(1+z^{-1})^n(1+z)^n}{\sigma_\zeta^2(1+z^{-1})^n(1+z)^n + \sigma_\varepsilon^2(1-z^{-1})^n(1-z)^n} \\ &= \frac{1}{1 + \lambda\left(\mathrm{i}\dfrac{1-z}{1+z}\right)^{2n}} \quad \text{with} \quad \lambda = \frac{\sigma_\varepsilon^2}{\sigma_\zeta^2}.\end{aligned} \tag{46}$$

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

$$\lambda = \{1/\tan(\omega_d)\}^{2n}, \tag{47}$$

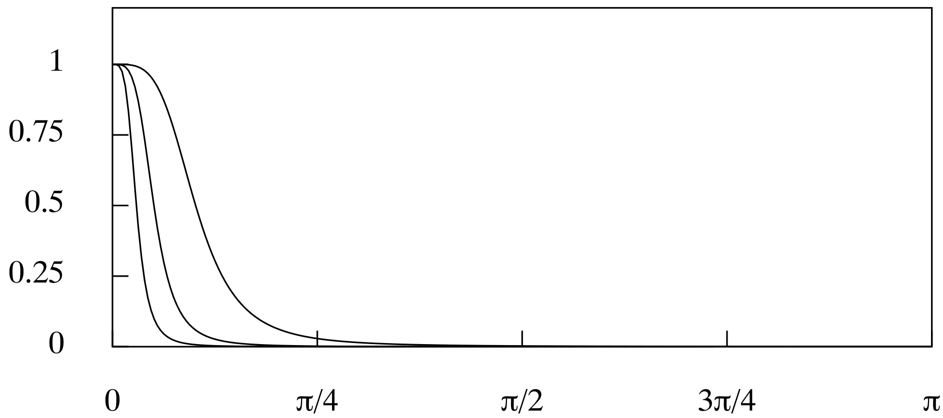

Figure 6. The gain of the Hodrick–Prescott lowpass filter with a smoothing parameter set to 100, 1,600 and 14,400.

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 bidirectional manner indicated by equation (40), 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 6 shows the square gain of the filter corresponding to these values. The innermost curve corresponds to $\lambda = 14{,}400$ and the outermost curve to $\lambda = 100$.

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 10, in the case of the U.K. consumption data.

It will be observed that an H–P filter with $\lambda = 1{,}600$, which defines the middle curve in Figure 6, will not be effective in isolating the low-frequency

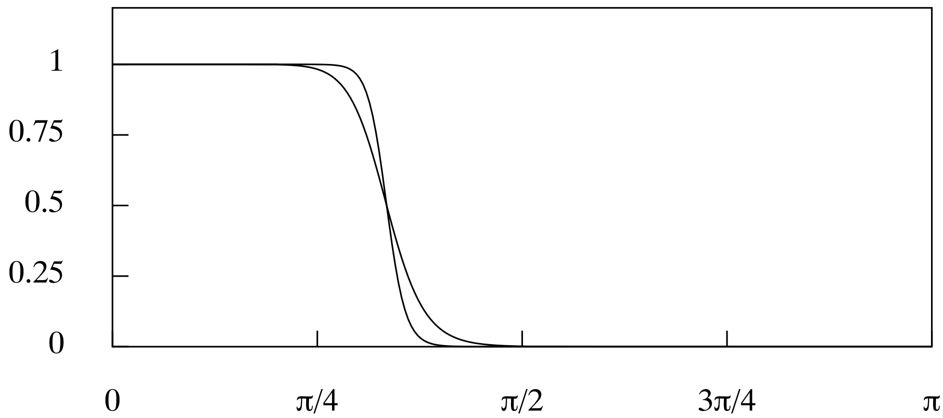

Figure 7. The squared gain of the lowpass Butterworth filters of orders $n = 6$ and $n = 12$ with a nominal cut-off point of $2\pi/3$ radians.

component of the quarterly consumption data of Figure 9, which lies in the interval $[0, \pi/8]$. The curve will cut through the low-frequency spectral structure that is represented in Figure 10; 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 7. In this case, there is a well-defined nominal cut-off frequency, which is at 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 analogue Butterworth filter is a commonplace of electrical engineering. The digital version has been described by Pollock (2000).

Wiener–Kolmogorov Filters for Finite Sequences

The classical Wiener–Kolmogorov theory can be adapted to finite data sequences generated by stationary stochastic processes.

Consider a data vector $y = [y_0, y_1, \ldots, y_{T-1},]'$ that has a signal component ξ and a noise component η:

$$y = \xi + \eta. \tag{48}$$

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

$$
\begin{aligned}
E(\xi) &= 0, \qquad D(\xi) = \Omega_\xi, \\
E(\eta) &= 0, \qquad D(\eta) = \Omega_\eta, \\
&\text{and} \quad C(\xi, \eta) = 0.
\end{aligned} \tag{49}
$$

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 $L_T = [e_1, e_2, \ldots, e_{T-1}, 0]$, which is the finite sample-version of the lag operator. This is obtained from the identity matrix $I_T = [e_0, e_1, e_2, \ldots, e_{T-1}]$ by deleting the leading column and by appending a zero vector to the end of the array. 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 $D(y) = \Omega_\xi + \Omega_\eta$.

We may begin by considering the determination of the vector of the T filter coefficients $\psi_{t.} = [\psi_{t,0}, \psi_{t,1}, \ldots, \psi_{t,T-1}]$ that determine x_t, which is the tth element of the filtered vector $x = [x_0, x_1, \ldots, x_{T-1}]'$. The estimate of ξ_t is

$$
x_t = \sum_{j=-t}^{t-T+1} \psi_{t,j} y_{t-j}, \tag{50}
$$

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

$$
\begin{aligned}
0 &= E\big\{y_{t-k}(\xi_t - x_t)\big\} \\
&= E(y_{t-k}\xi_t) - \sum_{j=-t}^{t-T+1} \psi_{t,j} E(y_{t-k} y_{t-j}) \\
&= \gamma_k^{y\xi} - \sum_{j=-t}^{t-T+1} \psi_{t,j} \gamma_{j-k}^{yy}.
\end{aligned} \tag{51}
$$

Equation (51) can be rendered also in a matrix format. By running from $k = -t$ to $k = T - t - 1$, we get the following system:

$$
\begin{bmatrix} \gamma_t^{\xi\xi} \\ \gamma_{t+1}^{\xi\xi} \\ \vdots \\ \gamma_{T-1-t}^{\xi\xi} \end{bmatrix}
=
\begin{bmatrix} \gamma_0^{yy} & \gamma_1^{yy} & \cdots & \gamma_{T-1}^{yy} \\ \gamma_1^{yy} & \gamma_0^{yy} & \cdots & \gamma_{T-2}^{yy} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1}^{yy} & \gamma_{T-2}^{yy} & \cdots & \gamma_0^{yy} \end{bmatrix}
\begin{bmatrix} \psi_{t,0} \\ \psi_{t,1} \\ \vdots \\ \psi_{t,T-1} \end{bmatrix}. \tag{52}
$$

Here, on the LHS, we have set $\gamma_j^{y\xi} = \gamma_j^{\xi\xi}$ in accordance with (33).

This equation above can be written in summary notation as $\Omega_\xi e_t = \Omega_y \psi'_{t.}$, where e_t is a vector of order T containing a single unit preceded by t zeros and followed by $T-1-t$ zeros. The coefficient vector $\psi_{t.}$ is given by

$$\psi_{t.} = e'_t \Omega_\xi \Omega_y^{-1} = e'_t \Omega_\xi (\Omega_\xi + \Omega_\eta)^{-1}, \tag{53}$$

and the estimate of ξ_t is $x_t = \psi_{t.}y$. Given the data $y = [y_0, y_1 \ldots, y_{T-1}]'$, the estimate of the complete vector $\xi = [\xi_0, \xi_1, \ldots, \xi_{T-1}]'$ of the corresponding signal elements would be

$$x = \Omega_\xi \Omega_y^{-1} y = \Omega_\xi (\Omega_\xi + \Omega_\eta)^{-1} y. \tag{54}$$

The Estimates as Conditional Expectations

The linear estimates of (54) have the status of conditional expectations, when the vectors ξ and y are normally distributed. As such, they are, unequivocally, the optimal minimun-mean-square-error predictors of the signal and the noise components:

$$\begin{aligned} E(\xi|y) &= E(\xi) + C(\xi, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1} y = x, \end{aligned} \tag{55}$$

$$\begin{aligned} E(\eta|y) &= E(\eta) + C(\eta, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1} y = h. \end{aligned} \tag{56}$$

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

$$\begin{aligned} D(\xi|y) &= D(\xi) - C(\xi, y)D^{-1}(y)C(y, \xi) \\ &= \Omega_\xi - \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\xi, \end{aligned} \tag{57}$$

$$\begin{aligned} D(\eta|y) &= D(\eta) - C(\eta, y)D^{-1}(y)C(y, \eta), \\ &= \Omega_\eta - \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\eta. \end{aligned} \tag{58}$$

The Least-Squares Derivation of the Estimates

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

$$\text{Minimise} \quad S(\xi, \eta) = \xi'\Omega_\xi^{-1}\xi + \eta'\Omega_\eta^{-1}\eta \quad \text{subject to} \quad \xi + \eta = y. \tag{59}$$

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.

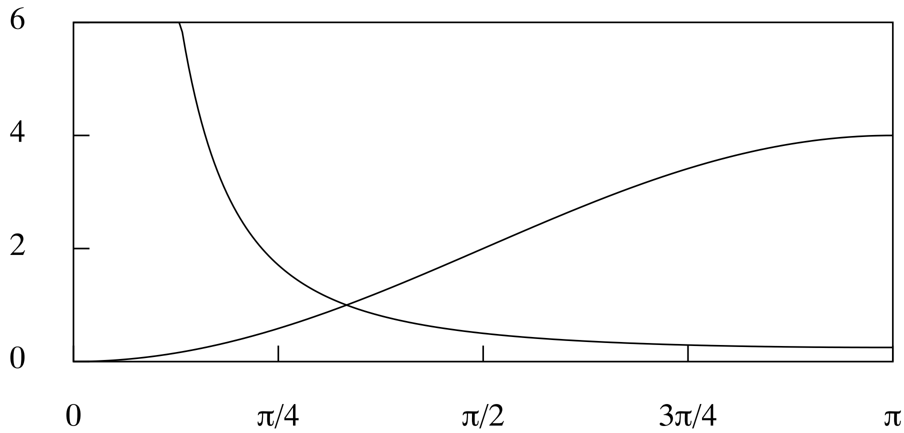

Figure 8. The squared gain of the difference operator, which has a zero at zero frequency, and the squared gain of the summation operator, which is unbounded at zero frequency.

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 $\Omega^{-1}(y-x) = \Omega_\xi^{-1}x$, which, on pre multiplication by Ω_η, can be written as $y = x + \Omega\Omega_\xi^{-1}x = (\Omega_\xi + \Omega_\eta)\Omega_\xi^{-1}x$. Therefore, the solution for x is

$$x = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y. \tag{60}$$

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

$$h = \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y \quad \text{and} \quad x = y - \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y. \tag{61}$$

The filter matrices $\Psi_\xi = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}$ and $\Psi_\eta = \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}$ of (60) and (61) are the matrix analogues of the z-transforms displayed in equations (38) and (39).

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

$$(\Omega_\xi + \Omega_\eta)b = y \tag{62}$$

for the value of b. Thereafter, one can generate

$$x = \Omega_\xi b \quad \text{and} \quad h = \Omega_\eta b. \tag{63}$$

If Ω_ξ and Ω_η correspond to the narrow-band dispersion matrices of moving-average processes, then the solution to equation (62) may be found via a Cholesky factorisation that sets $\Omega_\xi + \Omega_\eta = GG'$, where G is a lower-triangular matrix with a limited number of nonzero bands. The system $GG'b = y$ may be cast in the form of $Gp = y$ and solved for p. Then, $G'b = p$ can be solved for b.

The Difference and Summation Operators

A simple expedient for eliminating the trend from the data sequence $y(t) = \{y_t; t = 0, \pm 1, \pm 2, \ldots\}$ is to replace the sequence by its differenced version $y(t) - y(t-1)$ or by its twice differenced version $y(t) - 2y(t-1) + y(t-2)$. Differences of higher orders are rare. The z-transform of the difference is $(1-z)y(z) = y(z) - zy(z)$. On defining the operator $\nabla(z) = 1 - z$, the second differences can be expressed as $\nabla^2(z)y(t) = (1 - 2z + z^2)y(z)$.

The inverse of the difference operator is the summation operator

$$\Sigma(z) = (1-z)^{-1} = \{1 + z + z^2 + \cdots\}. \tag{64}$$

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

$$\Sigma^d(z) = \frac{1}{(1-z)^d} = 1 + dz + \frac{d(d+1)}{2!}z^2 + \frac{d(d+1)(d+2)}{3!}z^3 + \cdots. \tag{65}$$

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 8, 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.

In dealing with a finite sequence, it is appropriate to consider a matrix version of the difference operator. In the case of a sample of T elements comprised by the vector $y = [y_0, y_1, \ldots, y_{T-1}]'$, it is appropriate to use the matrix difference operator $\nabla(L_T) = I_T - L_T$, which is obtained by replacing z within $\nabla(z) = 1 - z$ by the matrix argument $L_T = [e_1, e_2, \ldots, e_{T-1}, 0]$, which is obtained from the identity matrix $I_T = [e_0, e_1, e_2, \ldots, e_{T-1}]$ by deleting the leading column and by appending a zero vector to the end of the array.

Examples of the first-order and second-order matrix difference operators are as follows:

$$\nabla_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad \nabla_4^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix}. \tag{66}$$

The corresponding inverse matrices are

$$\Sigma_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \Sigma_4^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ 4 & 3 & 2 & 1 \end{bmatrix}. \tag{67}$$

It will be seen that the elements of the leading vectors of these matrices are the coefficients associated with the expansion of $\Sigma^d(z)$ of (65) for the cases of $d = 1$ and $d = 2$. The same will be true for higher orders of d.

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:

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 6 & 3 & 1 \\ 10 & 6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -2 & -1 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}. \tag{68}$$

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

$$\nabla_T^p = (I - L_T)^p. \tag{69}$$

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

$$\nabla_T^p y = \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} y = \begin{bmatrix} g_* \\ g \end{bmatrix}; \tag{70}$$

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

$$\begin{bmatrix} S_* & S \end{bmatrix} \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} = S_* Q_*' + SQ' = I_T, \tag{71}$$

and that

$$\begin{bmatrix} Q_*' \\ Q' \end{bmatrix} \begin{bmatrix} S_* & S \end{bmatrix} = \begin{bmatrix} Q_*' S_* & Q_*' S \\ Q' S_* & Q' S \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_{T-p} \end{bmatrix}. \tag{72}$$

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

$$y = S_* g_* + Sg. \tag{73}$$

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

$$(y-x)'(y-x) = (y - S_* r_*)'(y - S_* r_*) \tag{74}$$

with respect to r_*. The resulting values are

$$r_* = (S_*' S_*)^{-1} S_*' y \quad \text{and} \quad x = S_*(S_*' S_*)^{-1} S_*' y. \tag{75}$$

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

$$S_*(S_*' S_*)^{-1} S_*' = I - Q(Q'Q)^{-1} Q'. \tag{76}$$

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

$$\begin{aligned} Z(Z'Z)^{-1}Z' &= \begin{bmatrix} Q & S_* \end{bmatrix} \begin{bmatrix} (Q'Q)^{-1} & 0 \\ 0 & (S_*' S)^{-1} \end{bmatrix} \begin{bmatrix} Q' \\ S_*' \end{bmatrix} \\ &= Q(Q'Q)^{-1}Q' + S_*(S_*' S_*)^{-1} S_*'. \end{aligned} \tag{77}$$

The result of (76) follows from the fact that $Z(Z'Z)^{-1}Z' = Z(Z^{-1}Z'^{-1})Z' = I$. It follows from (76) that the vector of the ordinates of the polynomial regression is also given by

$$x = y - Q(Q'Q)^{-1}Q'y. \tag{78}$$

Polynomial Regression and Trend Extraction

The use of polynomial regression in a preliminary detrending of the data is an essential part of a strategy for determining an appropriate representation

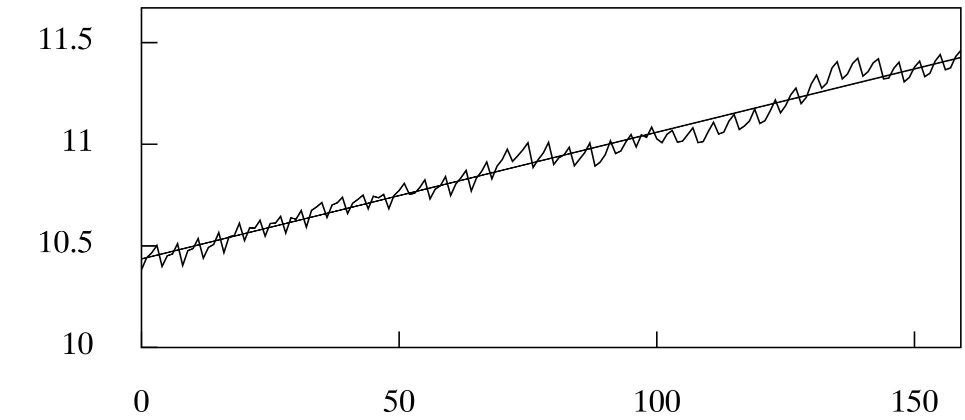

Figure 9. The quarterly series of the logarithms of consumption in the U.K., for the years 1955 to 1994, together with a linear trend interpolated by least-squares regression.

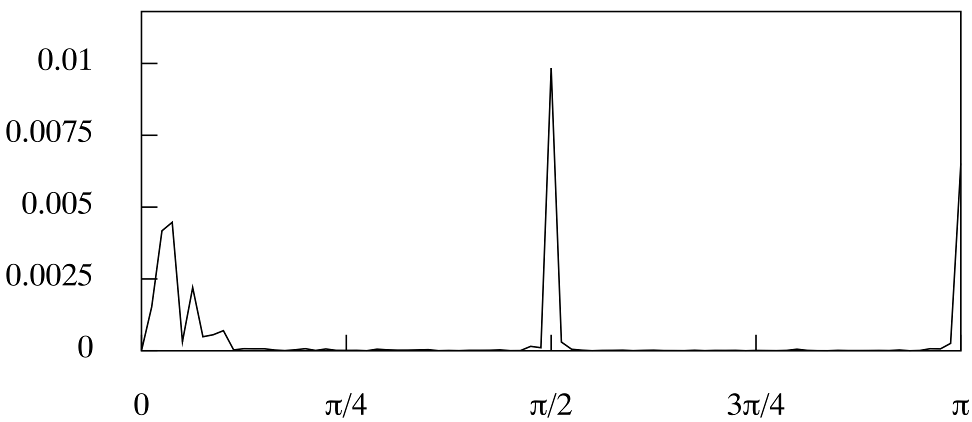

Figure 10. The periodogram of the residual sequence obtained from the linear detrending of the logarithmic consumption data.

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 9 and 10. Figure 9 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 exponential growth, against which the fluctuations of consumption can be measured. The periodogram of the residual sequence in plotted in Figure 10. 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

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

$$\nabla_5^2 = \begin{bmatrix} Q'_* \\ Q' \end{bmatrix} = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ \hline 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{array}\right]. \tag{79}$$

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

$$\begin{aligned} Q'y &= Q'\xi + Q'\eta \\ &= \delta + \kappa = g. \end{aligned} \tag{80}$$

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

$$\begin{aligned} E(\delta) &= 0, \qquad D(\delta) = \Omega_\delta = Q'D(\xi)Q, \\ E(\kappa) &= 0, \qquad D(\kappa) = \Omega_\kappa = Q'D(\eta)Q. \end{aligned} \tag{81}$$

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

$$\begin{aligned} h = E(\eta|g) &= E(\eta) + C(\eta, g)D^{-1}(g)\{g - E(g)\} \\ &= C(\eta, g)D^{-1}(g)g, \end{aligned} \tag{82}$$

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

$$D(g) = \Omega_\delta + Q'\Omega_\eta Q \quad \text{and} \quad C(\eta, g) = \Omega_\eta Q. \tag{83}$$

Putting these details into (82) gives the following estimate of η:

$$h = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{84}$$

Putting this into the equation

$$x = E(\xi|g) = y - E(\eta|g) = y - h \tag{85}$$

gives

$$x = y - \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{86}$$

The Least-Squares Derivation of the Filter

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

$$\text{Minimise} \quad (y - \xi)'\Omega_\eta^{-1}(y - \xi) + \xi' Q\Omega_\delta^{-1}Q'\xi. \tag{87}$$

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

$$\Omega_\eta^{-1}(y - x) = -Q\Omega_\delta^{-1}Q'x = Q\Omega_\delta^{-1}d, \tag{88}$$

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

$$Q'(y - x) = Q'y - d = Q'\Omega_\eta Q\Omega_\delta^{-1}d, \tag{89}$$

whence

$$\begin{aligned} Q'y &= d + Q'\Omega_\eta Q\Omega_\delta^{-1}d \\ &= (\Omega_\delta + Q'\Omega_\eta Q)\Omega_\delta^{-1}d, \end{aligned} \tag{90}$$

which gives

$$\Omega_\delta^{-1}d = (\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{91}$$

Putting this into

$$x = y - \Omega_\eta Q\Omega_\delta^{-1}d, \tag{92}$$

which comes from premultiplying (88) by Ω_η, gives

$$x = y - \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{93}$$

One should observe that

$$\Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'e, \tag{94}$$

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 Leser (H–P) Filter and the Butterworth Filter

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

$$D(\eta) = \Omega_\eta = \sigma_\eta^2 I, \quad D(\delta) = \Omega_\delta = \sigma_\delta^2 I \quad \text{and} \quad \lambda = \frac{\sigma_\eta^2}{\sigma_\delta^2} \tag{95}$$

within (93) to give

$$x = y - Q(\lambda^{-1}I + Q'Q)^{-1}Q'y \tag{96}$$

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 (78). The matrix expression $\Psi = I - Q(\lambda^{-1}I + Q'Q)^{-1}Q'$ for the filter can be compared to the polynomial expression $\psi^c(z) = 1 - \psi(z)$ of the classical formulation, which entails the z-transform from (45).

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

$$x = y - \lambda\Sigma Q(M + \lambda Q'\Sigma Q)^{-1}Q'y. \tag{97}$$

Here, the matrices

$$\Sigma = \{2I_T - (L_T + L_T^{-1})\}^{n-2} \quad \text{and} \quad M = \{2I_T + (L_T + L_T^{-1})\}^n \tag{98}$$

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

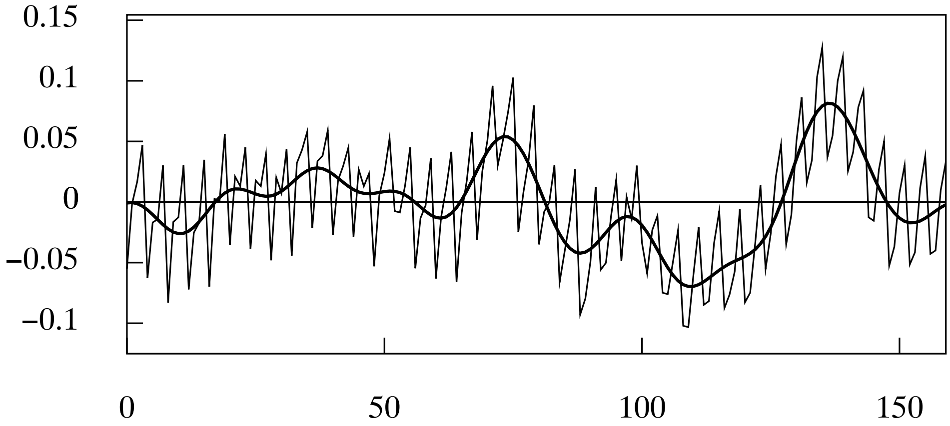

Figure 11. The residual sequence from fitting a linear trend to the logarithmic consumption data with an interpolated line representing the business cycle, obtained by the frequency-domain method.

z by L_T and z^{-1} by L_T^{-1}. Observe that the equalities no longer hold after the replacements. However, it can be verified that

$$Q'\Sigma Q = \{2I_T - (L_T + L_T^{-1})\}^n. \tag{99}$$

Filtering in the Frequency Domain

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

$$\begin{aligned} &\Omega_\xi^\circ = \bar{U}\gamma_\xi(D)U, \quad \Omega_\eta^\circ = \bar{U}\gamma_\eta(D)U \quad \text{and} \\ &\Omega^\circ = \Omega_\xi^\circ + \Omega_\eta^\circ = \bar{U}\{\gamma_\xi(D) + \gamma_\eta(D)\}U, \end{aligned} \tag{100}$$

wherein the diagonal matrices $\gamma_\xi(D)$ and $\gamma_\eta(D)$ contain the ordinates of the spectral density functions of the component processes. Accounts of the algebra of circulant matrices have been provided by Pollock (1999 and 2002). See, also, Gray (2002).

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 $W^{jt} = \exp(-i2\pi tj/T)$, and $\bar{U}$ is its conjugate transpose. Also, $D = \text{diag}\{1, W, W^2, \ldots, W^{T-1}\}$, 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 (55) and (56) by their circulant counterparts, we derive the following formulae:

$$x = \bar{U}\gamma_\xi(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}Uy = P_\xi y, \tag{101}$$

$$h = \bar{U}\gamma_\eta(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}Uy = P_\eta y. \tag{102}$$

Similar replacements within the formulae (57) and (58) 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 example of the method of frequency filtering is provided by Figure 11, which shows the effect applying a filter with a sharp cut-off at the frequency value of $\pi/8$ radians per period to the residual sequence obtained from a linear detrending of the quarterly logarithmic consumption data of the U.K.

This cut-off frequency has been chosen in reference to the periodogram of the residual sequence, which is in Figure 10. This shows that the low-frequency structure of the data falls in the interval $[0, \pi/8]$. Apart from the prominent spike at the season frequency of $\pi/2$ and the smaller seasonal spike at the frequency of π, the remainder of the peridogram is characterised by wide spectral deadspaces.

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 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

$$x = S_*d_* + Sd \quad \text{where} \quad d_* = (S_*'S_*)^{-1}S_*'(y - Sd) \tag{103}$$

and

$$h = S_*k_* + Sk \quad \text{where} \quad k_* = -(S_*'S_*)^{-1}S_*'Sk. \tag{104}$$

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

$$\begin{aligned}(y - x)'(y - x) &= (y - S_*d_* - Sd)'(y - S_*d_* - Sd)\\ &= (S_*k_* + Sk)'(S_*k_* + Sk) = h'h.\end{aligned} \tag{105}$$

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

$$S_*' = \begin{bmatrix} 1 & 2 & \ldots & T-1 & T \\ 0 & 1 & \ldots & T-2 & T-1 \end{bmatrix} \tag{106}$$

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

$$(107)\qquad \begin{aligned}\sum_{t=1}^{T} t^2 &= \frac{1}{6}T(T+1)(2T+1) \quad \text{and}\\ \sum_{t=1}^{T} t(t-1) &= \frac{1}{6}T(T+1)(2T+1) - \frac{1}{2}T(T+1).\end{aligned}$$

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

A fuller account of the implementation of the frequency filter has been provided by Pollock (2009).

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

$$(108)\qquad b = \frac{\sum y_t^2 - (\sum y_t)^2/T}{\sum t^2 - (\sum t)^2/T}.$$

The formulae

$$(109)\qquad \sum_{t=0}^{T-1} t^2 = \frac{1}{6}(T-1)T(2T-1) \quad \text{and} \quad \sum_{t=0}^{T-1} t = \frac{T(T-1)}{2}$$

are combined to provide a convenient means of calculating the denominator of the formula of (108):

$$(110)\qquad \sum_{t=0}^{T-1} t^2 - \frac{(\sum_{t=0}^{T-1} t)^2}{T} = \frac{(T-1)T(T+1)}{12}.$$

Another means of calculating the low-frequency trajectory of the data via the frequency domain mimics the method of equation (93) 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 (93), 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

$$\begin{aligned} N(z) &= z^{-1} - 2 + z = z^{-1}(1-z)^2 \\ &= z^{-1}\nabla^2(z). \end{aligned} \tag{111}$$

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

$$N(L_T) = N_T = L_T - 2I_T + L'_T. \tag{112}$$

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'_0N_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

$$N_5 = \begin{bmatrix} Q'_{-1} \\ Q' \\ Q_{+1} \end{bmatrix} = \left[\begin{array}{ccccc} -2 & 1 & 0 & 0 & 0 \\ \hline 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ \hline 0 & 0 & 0 & 1 & -2 \end{array}\right]. \tag{113}$$

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 that 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

$$f(\omega) = 2 - 2\cos(\omega). \tag{114}$$

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

$$V = \text{diag}\{v_0, v_1, \ldots, v_{T-1}\}, \tag{115}$$

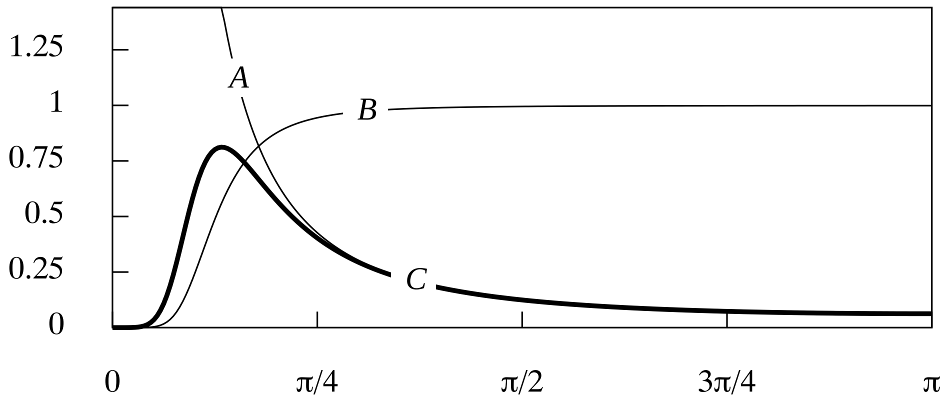

Figure 12. The pseudo-spectrum of a random walk, labelled A, together with the squared gain of the highpass Hodrick–Prescott filter with a smoothing parameter of $\lambda = 100$, labelled B. The curve labelled C represents the spectrum of the filtered process.

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

$$h = \bar{U}H\gamma = \bar{U}HUq. \tag{116}$$

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

$$x = y - h = y - \bar{U}HUq. \tag{117}$$

Business Cycles and Spurious Cycles

Econometricians continue to debate the question of how macroeconomic data sequences should be decomposed into their constituent components. These components are usually described as the trend, the cyclical component or the business cycle, the seasonal component and the irregular component.

For the original data, the decomposition is usually a multiplicative one and, for the logarithmic data, the corresponding decomposition is an additive one. The filters are usually applied to the logarithmic data, in which case, the sum of the estimated components should equal the logarithmic data.

In the case of the Wiener–Kolmogorov filters, and of the frequency-domain filters as well, the filter gain never exceeds unity. Therefore, every lowpass filter $\psi(z)$ is accompanied by a complementary highpass filter $\psi^c(z) = 1 - \psi(z)$. The two sequences resulting from these filters can be recombined to create the data sequence from which they have originated.

Such filters can be applied sequentially to create an additive decomposition of the data. First, the tend is extracted. Then, the cyclical component is extracted from the detrended data, Finally, the residue can be decomposed into the seasonal and the irregular components.

Within this context, the manner in which any component is defined and how it is extracted are liable to affect the definitions of all of the other components. In particular, variations in the definition of the trend will have substantial effects upon the representation of the business cycle.

It has been the contention of several authors, including Harvey and Jaeger (1993) and Cogley and Nason (1995), that the effect of using the Hodrick–Prescott filter to extract a trend from the data is to create or induce spurious cycles in the complementary component, which includes the cyclical component.

Others have declared that such an outcome is impossible. They point to the fact that, since the gains never exceeds unity, the filters cannot induce anything into the data, nor can they amplify anything that is already present. On this basis, it can be fairly asserted that, at least, the verbs *to create* and *to induce* have been miss-applied, and that the use of the adjective *spurious* is doubtful.

The analyses of Harvey and Jaeger and of Cogley and Nason have both depicted the effect of applying the Hodrick–Prescott filter to a theoretical random walk that is supported on a doubly-infinite set of integers. They show that the spectral density function of the filtered process possesses a peak in the low-frequency region that is based on a broad range of frequencies. This seems to suggest that there is cyclicality in the processed data, whereas the original random walk has no central tendency.

This analysis is illustrated in Figure 12. The curve labelled A is the pseudo spectrum of a first-order random walk. The curve labelled B is the squared modulus of the frequency response of the highpass, detrending, filter with a smoothing parameter of 100. The curve labelled C is the spectral density function of a detrended sequence which, in theory, would be derived by applying the filter to the random walk.

The fault of the Hodrick–Prescott filter may be that it allows elements of the data at certain frequencies to be transmitted when, ideally, they should be blocked. However, it seems that an analysis based on a doubly-infinite random walk is of doubtful validity.

The effects that are depicted in Figure 12 are due largely to the unbounded nature of the pseudo spectrum labelled A, and, as we have already declared, there is a zero probability that, at any given time, the value generated by the random walk will fall within a finite distance of the horizontal axis.

An alternative analysis of the filter can be achieved by examining the

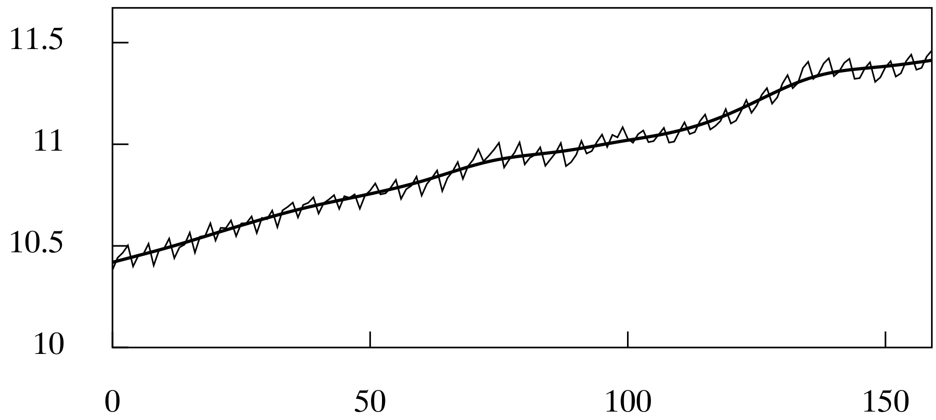

Figure 13. the quarterly logarithmic consumption data together with a trend interpolated by the lowpass Hodrick–Prescott filter with the smoothing parameter set to $\lambda = 1,600$.

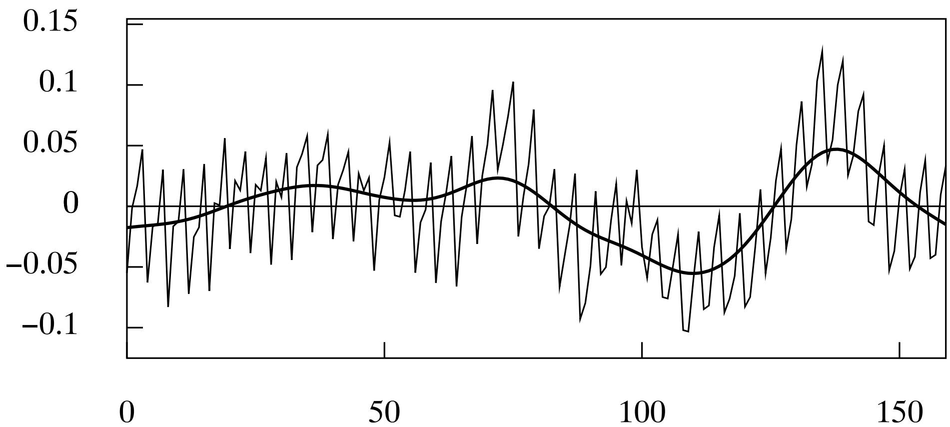

Figure 14. The residual sequence obtained by extracting a linear trend from the logarithmic consumption data, together with a low-frequency trajectory that has been obtained via the lowpass Hodrick–Prescott filter.

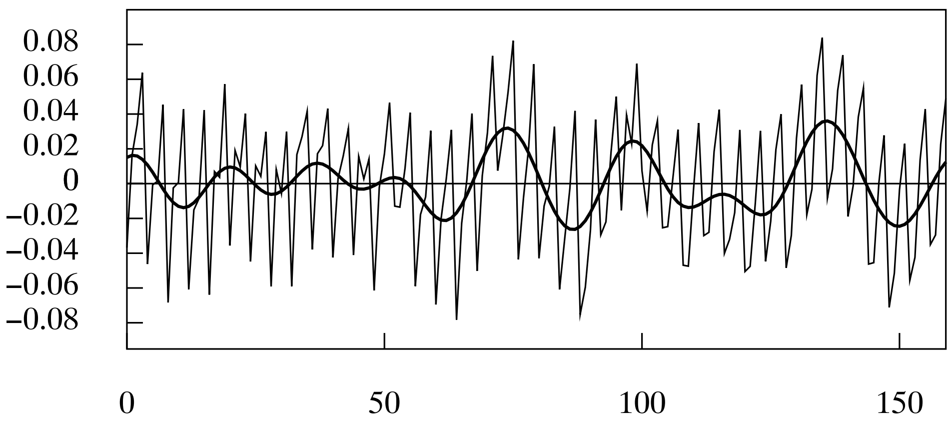

Figure 15. The residual sequence obtained by using the Hodrick–Prescott filter to extract the trend, together with a fluctuating component obtained by subjecting the sequence to a lowpass frequency-domain filter with a cut-off point at $\pi/8$ radians.

effects of its finite-sample version upon a sequence that has been detrended by the interpolation of a linear function, according to the ordinary least-squares criterion.

A linear function in logarithmic data corresponds to a trajectory in the original data of a process of constant exponential growth. Over a limited period, characterised by normal economic activity, this may well provide an appropriate benchmark against which to measure economic fluctuations.

It also follows, in consequence of equation (94), that the highpass filter will generate the same output from the linearly detrended data as from the original data. Therefore, in characterising the effects of the filter, it is reasonable to compare the linearly detrended data with the output of the filter.

Figure 13 shows the quarterly logarithmic consumption data together with a trend interpolated by the lowpass Hodrick–Prescott filter with the smoothing parameter set to $\lambda = 1{,}600$.

Figure 14 shows the residual sequence obtained by extracting a linear trend from the logarithmic consumption data. This sequence has also been shown in Figure 11, where it has been interpolated by a smooth cyclical trajectory obtained via a lowpass frequency-domain filter. Figure 14 also shows a low frequency trajectory that has been obtained by subjecting the residual sequence to the lowpass Hodrick–Prescott Filter with a smoothing parameter of 1,600,

The residual sequence from the latter process is also the residual sequence that would be obtained by the application of the highpass Hodrick–Prescott filter either to the original data or to the linearly detrended data. This sequence, which is shown in Figure 15, is beset by substantial seasonal fluctuations.

One way of removing the seasonal fluctuations is to subject the sequence to a lowpass frequency-domain filter with a cut-off at $\pi/8$ radians, which is the value that has already been determined from the inspection of the peridogram of Figure 10. This filtering operation has given rise to the fluctuating trajectory that is also to be found in Figure 15.

Now, a comparison can be made between the business cycle trajectory of Figure 11, which has been determined via a linear detrending of the logarithmic data, and the trajectory of Figure 15, which has been determined via the Hodrick–Prescott filter. Whereas the same essential fluctuations are present in both trajectories, it is apparent that the more flexible detrending of the Hodrick–Prescott filter has served to reduce and to regularise their amplitudes.

This effect is readily explained by reference to the least-squares criterion by which the Hodrick–Prescott filter can be derived. According to the criterion, larger deviations from the horizontal axis are penalised more than are smaller deviations. The effect is manifest in the comparison of Figures 11 and 15. It might be said, in summary, that the Hodrick–Prescott filter is liable to lend a spurious regularity to the fluctuations.

References

Cogley, T., and J.M. Nason, (1995), Effects of the Hodrick–Prescott Filter on Trend and Difference Stationary Time Series, Implications for Business Cycle

Research, *Journal of Economic Dynamics and Control,* 19, 253–278.

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

Hall, H.S., and S.R. Knight, (1899), *Higher Algebra,* Macmillan and Co., London.

Harvey, A.C., and A. Jaeger, (1993), Detrending, Stylised Facts and the Business Cycle, *Journal of Applied Econometrics,* 8, 231–247.

Hodrick, R.J., and E.C. Prescott, (1980), *Postwar U.S. Business Cycles: An Empirical Investigation,* Working Paper, Carnegie–Mellon University, Pittsburgh, Pennsylvania.

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

Jolly, L.B.W., (1961), *Summation of Series: Second Revised Edition,* Dover Publications: New York.

Jury, E.I., (1964), *Theory and Applications of the z-Transform Method,* John Wiley and Sons, New York.

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

Leser, C.E.V., (1961), A Simple Method of Trend Construction, *Journal of the Royal Statistical Society, Series B,* 23, 91–107.

Nyquist, H., (1928), Certain Topics in Telegraph Transmission Theory, *AIEE Transactions, Series B,* 617–644.

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

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

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

Pollock, D.S.G., (2009), Realisations of Finite-sample Frequency-selective Filters, *Journal of Statistical Planning and Inference,* 139, 1541–1558.

Shannon, C.E., (1949a), Communication in the Presence of Noise, *Proceedings of the Institute of Radio Engineers,* 37, 10–21. Reprinted in 1998, *Proceedings of the IEEE,* 86, 447–457.

Shannon, C.E., (1949b), (reprinted 1998), *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, Illinois.

Whittaker, E.T., (1923), On a New Method of Graduations, *Proceedings of the Edinburgh Mathematical Society,* 41, 63-75.

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.