FILTERS FOR ECONOMETRIC DATA

Wiener–Kolmogorov Filtering of Stationary Sequences

The classical theory of linear filtering was formulated independently by Norbert Wiener (1941) and Andrei Nikolaevich Kolmogorov (1941) during the Second World War. They were both considering the problem of how to target radar-assisted anti-aircraft guns on incoming enemy aircraft.

The theory has found widespread application in analog and digital signal processing and in telecommunications in general. Also, it has provided a basic technique for the enhancement of recorded music.

The classical theory assumes that the data sequences are generated by stationary stochastic processes and that these are of sufficient length to justify the assumption that they constitute doubly-infinite sequences.

For econometrics, the theory must to be adapted to cater to short trended sequences. Then, Wiener–Kolmogorov filters can used to extract trends from economic data sequences and for generating seasonally adjusted data.

Consider a vector y with a signal component ξ and a noise component η:

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

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

$$\begin{aligned} E(\xi) &= 0, \qquad D(\xi) = \Omega_\xi, \\ E(\eta) &= 0, \qquad D(\eta) = \Omega_\eta, \\ \text{and} \quad & C(\xi, \eta) = 0. \end{aligned} \tag{2}$$

A consequence of the independence of ξ and η is that

$$D(y) = \Omega_\xi + \Omega_\eta \quad \text{and} \quad C(\xi, y) = D(\xi) = \Omega_\xi. \tag{3}$$

The signal component is estimated by a linear transformation $x = \Psi_x y$ of the data vector that suppresses the noise component. Usually, the signal comprises low-frequency elements and the noise comprises elements of higher frequencies.

The Minimum Mean-Squared Error Estimator

The principle of linear minimum mean-squared error estimation indicates that the error $\xi - x$ in representing ξ by x should be uncorrelated with the data in y:

$$\begin{aligned} 0 = C(\xi - x, y) &= C(\xi, y) - C(x, y) \\ &= C(\xi, y) - \Psi_x C(y, y) \\ &= \Omega_\xi - \Psi_x(\Omega_\xi + \Omega_\eta). \end{aligned} \tag{4}$$

The solution is $\Psi_x = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}$ and the estimate of the signal is

$$x = \Psi_x y = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1} y. \tag{5}$$

The corresponding estimate of the noise component η is

$$\begin{aligned} h = \Psi_h y &= \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1} y \\ &= \{I - \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}\} y. \end{aligned} \tag{6}$$

It will be observed that $\Psi_\xi + \Psi_\eta = I$ and, therefore, that $x + h = y$.

Conditional Expectations

In deriving the estimator, we might have used the formula for conditional expectations. In the case of two linearly related scalar random variables ξ and y, the conditional expectation of ξ given y is

$$E(\xi|y) = E(\xi) + \frac{C(\xi, y)}{V(y)}\{y - E(y)\} \tag{7}$$

In the case of two vector quantities, this becomes

$$E(\xi|y) = E(\xi) + C(\xi, y)D^{-1}(y)\{y - E(y)\} \tag{8}$$

By setting

$$C(\xi, y) = \Omega_\xi \quad \text{and} \quad D(y) = \Omega_\xi + \Omega_\eta$$

as in (3), and by setting $E(\xi) = E(y) = 0$, we get the expression that is to be found under (5):

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

The Difference Operator and Polynomial Regression

The lag operator L, which is commonly defined in respect of a doubly-infinite sequence $x(t) = \{x_t; t = 0 \pm 1, \pm 2, \ldots\}$, has the effect that $Lx(t) = x(t-1)$.

The (backwards) difference operator $\nabla = 1 - L$ has the effect that $\nabla x(t) = x(t) - x(t-1)$. It serves to reduce a constant function to zero and to reduce a linear function to a constant. The second-order or twofold difference operator

$$\nabla^2 = 1 - 2L + L^2$$

is effective in reducing a linear function to zero.

A difference operator ∇^d of order d is commonly employed in the context of an ARIMA(p, d, q) model to reduce the data to stationarity. Then, the differenced data can be modelled by an ARMA(p, q) process. In such circumstances, the difference operator takes the form of a matrix transformation.

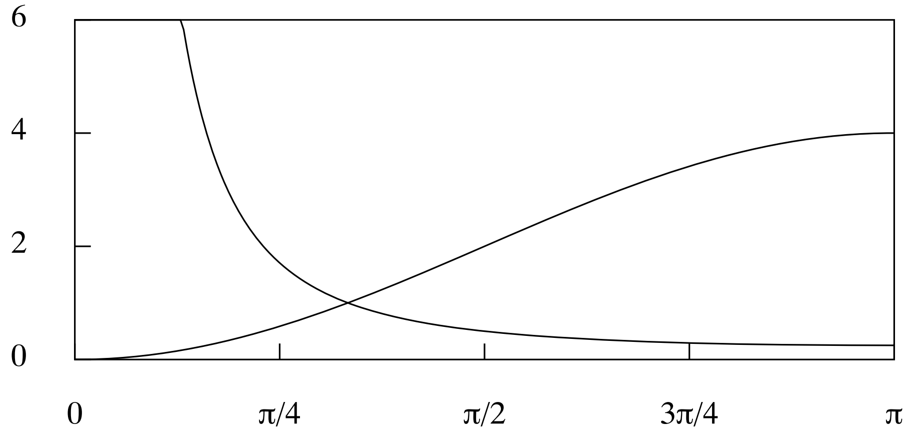

Figure 1. The squared gain of the difference operator, which has a zero at zero frequency, and the squared gain of the summation operator, which is unbounded at zero frequency.

The Matrix Difference Operator

The matrix analogue of the second-order difference operator in the case of $T = 5$, for example, is given by

$$\nabla_5^2 = \begin{bmatrix} Q'_* \\ Q' \end{bmatrix} = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ \hline 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{array}\right]. \tag{9}$$

The first two rows, which do not produce true differences, are liable to be discarded.

The difference operator nullifies data elements at zero frequency and it severely attenuates those at the adjacent frequencies. This is a disadvantage when the low frequency elements are of primary interest. Another way of detrending the data is to fit a polynomial trend by least-squares regression and to take the residual sequence as the detrended data.

Polynomial Regression

Using the matrix Q' defined above, we can represent the vector of the ordinates of a linear trend line interpolated through the data sequence as

$$x = y - Q(Q'Q)^{-1}Q'y. \tag{10}$$

The vector of the residuals is

$$e = Q(Q'Q)^{-1}Q'y. \tag{11}$$

Observe that this vector contains exactly the same information as the differenced vector $g = Q'y$. However, whereas the low-frequency structure of

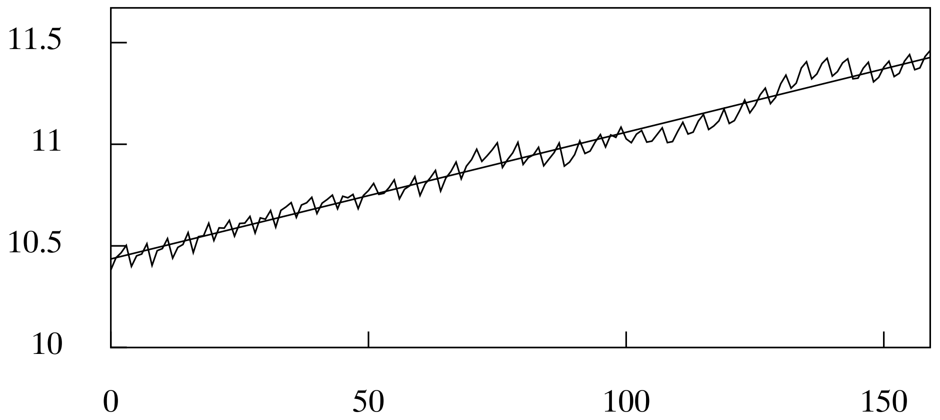

Figure 2. The quarterly series of the logarithms of consumption in the U.K., for the years 1955 to 1994, together with a linear trend interpolated by least-squares regression.

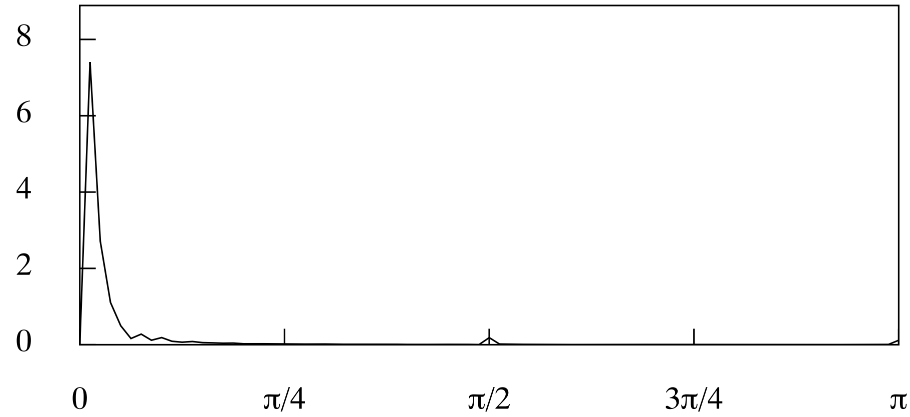

Figure 3. The periodogram of the trended logarithmic data.

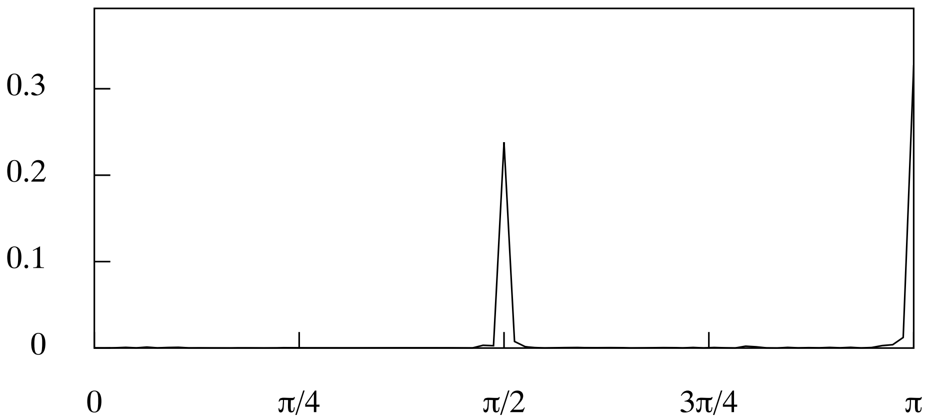

Figure 4. The periodogram of the differenced logarithmic consumpption data.

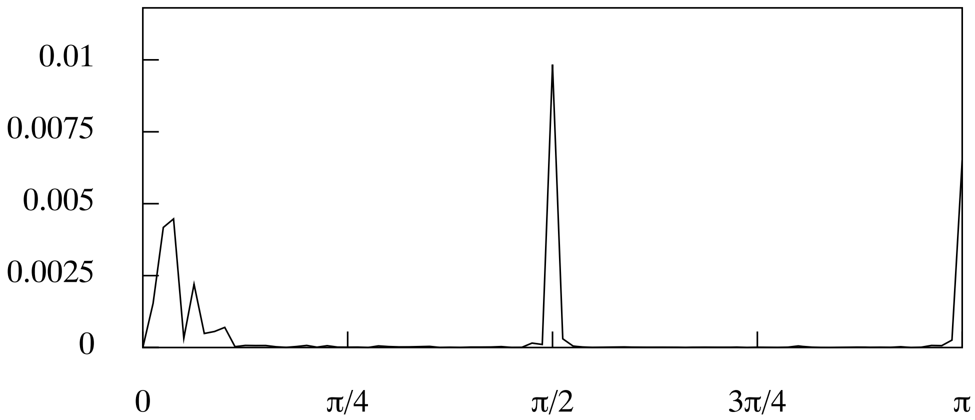

Figure 5. The periodogram of the residual sequence obtained from the linear detrending of the logarithmic consumption data.

the data in invisible in the periodogram of the latter, it is entirely visible in the periodogram of the residuals.

Filters for Short Trended Sequences

Applying Q' to the equation $y = \xi + \eta$, representing the trended data, gives

$$\begin{aligned} Q'y &= Q'\xi + Q'\eta \\ &= \delta + \kappa = g. \end{aligned} \tag{12}$$

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

$$\begin{aligned} E(\delta) &= 0, \qquad D(\delta) = \Omega_\delta = Q'D(\xi)Q, \\ E(\kappa) &= 0, \qquad D(\kappa) = \Omega_\kappa = Q'D(\eta)Q. \end{aligned} \tag{13}$$

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

$$\begin{aligned} h = E(\eta|g) &= E(\eta) + C(\eta, g)D^{-1}(g)\{g - E(g)\} \\ &= C(\eta, g)D^{-1}(g)g, \end{aligned} \tag{14}$$

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

$$D(g) = \Omega_\delta + Q'\Omega_\eta Q \quad \text{and} \quad C(\eta, g) = \Omega_\eta Q. \tag{15}$$

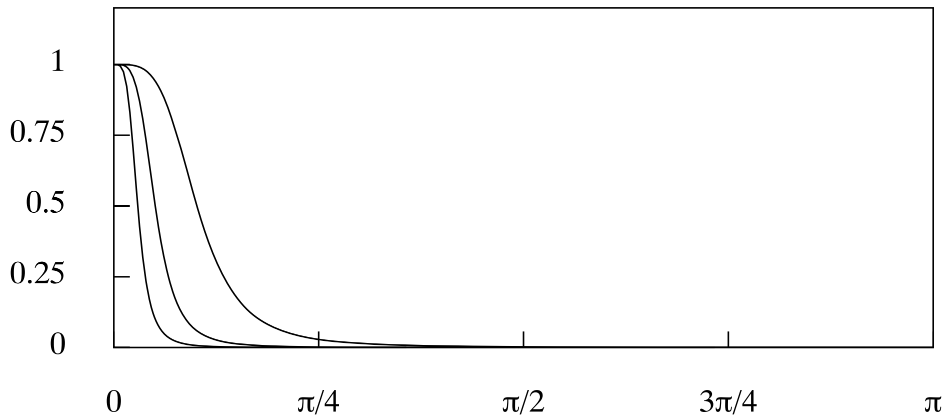

Figure 6. The gain of the Hodrick–Prescott lowpass filter with a smoothing parameter set to 100, 1,600 and 14,400.

Putting these details into (14) gives the following estimate of η:

$$h = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{16}$$

Putting this into the equation $x = y - h$ gives

$$x = y - \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{17}$$

The Leser (H–P) Filter

We now consider two specific cases of the Wiener–Kolmogorov filter. First, there is the Leser or Hodrick–Prescott (H–P) filter. This can be derived from a model that supposes that the signal is generated by an integrated (second-order) random walk and and that the noise is from a white-noise process.

The random walk process is reduced to a white-noise process $\delta(t)$ by taking twofold differences. Thus, $(1 - L)^2\xi(t) = \delta(t)$, and the corresponding equation for the sample is $Q'\xi = \delta$. Accordingly, the filter is derived by setting

$$D(\eta) = \Omega_\eta = \sigma_\eta^2 I, \quad D(\delta) = \Omega_\delta = \sigma_\delta^2 I \quad \text{and} \quad \lambda = \frac{\sigma_\eta^2}{\sigma_\delta^2} \tag{18}$$

within (17) to give

$$x = y - Q(\lambda^{-1}I + Q'Q)^{-1}Q'y. \tag{19}$$

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 (10):

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

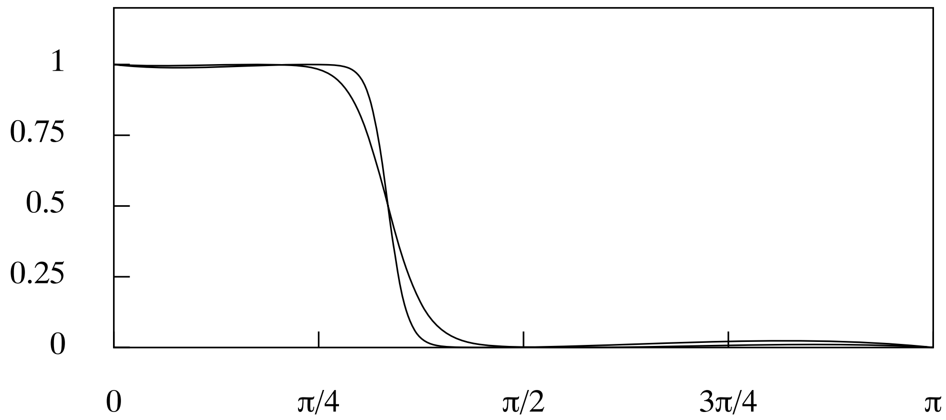

Figure 7. The gain of the lowpass Butterworth filters of orders $n = 6$ and $n = 12$ with a nominal cut-off point of $2\pi/3$ radians.

Figure 6 depicts the frequency response of the lowpass H–P filter for various values of the smoothing parameter λ. The innermost profile corresponds to the highest value of the parameter, and it represents a filter that transmits only the data elements of lowest frequency.

For all values of λ, the response of the H–P filter shows a gradual transition from the pass band, which corresponds to the frequencies that are transmitted by the filter, to the stop band, which corresponds to the frequencies that are impeded.

Often, there is a requirement for a more rapid transition as well as a need to control the location in frequency where the transitions occurs. These needs can be served by the Butterworth filter, which is more amenable to adjustment.

The Butterworth Filter

The Butterworth filter can be derived from an heuristic model in which the signal and the noise are generated by processes that are described, respectively, by the equations $(1-L)^2\xi(t) = (1+L)^n\zeta(t)$ and $(1-L)^2\eta(t) = (1-L)^n\varepsilon(t)$, where $\zeta(t)$ and $\eta(t)$ are mutually independent white-noise processes.

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

$$x = y - \lambda\Sigma Q(M + \lambda Q'\Sigma Q)^{-1}Q'y. \tag{20}$$

Here, the matrices are

$$\Sigma = \{2I_T - (L_T + L'_T)\}^{n-2} \quad \text{and} \quad M = \{2I_T + (L_T + L'_T)\}^n, \tag{21}$$

where L_T is a matrix of order T with units on the first subdiagonal; and it can be verified that

$$Q'\Sigma Q = \{2I_T - (L_T + L'_T)\}^n. \tag{22}$$

Figure 7 shows the frequency response of the Butterworth filter for various values of n and for a specific cut-off frequency, which is determined by the parameter λ. The greater the value of n, the more rapid is the transition from pass band to stop band.