8. FILTERING SHORT NONSTATIONARY SEQUENCES

Wiener-Kolmogorov Filtering of Short Stationary Sequences

Assume that the observations running from $t = 0$ to $t = T - 1$ are gathered in a vector

$$y = \xi + \eta, \tag{8.1}$$

where ξ is the signal component and η is the noise component. It may be assumed that these are from independent zero-mean Gaussian processes that are completely characterised by their first and second moments. These are

$$\begin{aligned} E(\xi) &= 0, \qquad D(\xi) = \Omega_\xi, \\ E(\eta) &= 0, \qquad D(\eta) = \Omega_\eta, \\ \text{and} \quad & C(\xi, \eta) = 0. \end{aligned} \tag{8.2}$$

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

The autocovariance or dispersion matrices, which have a Toeplitz structure, may be obtained by replacing the argument z within the relevant autocovariance generating functions by the matrix $L_T = [e_1, \ldots, e_{T-1}, 0]$.

Using L_T in place of z in the autocovariance function $\gamma(z)$ of the data process gives

$$D(y) = \Omega = \gamma_0 I_T + \sum_{\tau=1}^{T-1} \gamma_\tau (L_T^\tau + F_T^\tau), \tag{8.3}$$

where $F_T = L'_T$ is in place of z^{-1}. Since L_T and F_T are nilpotent of degree T, such that $L_T^q, F_T^q = 0$ when $q \geq T$, the index of summation has an upper limit of $T - 1$.

The optimal predictors of the signal and the noise components are their minimum-mean-square-error estimates

$$\begin{aligned} E(\xi|y) &= E(\xi) + C(\xi, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y = Z_\xi y = x, \end{aligned} \tag{8.4}$$

$$\begin{aligned} E(\eta|y) &= E(\eta) + C(\eta, y)D^{-1}(y)\{y - E(y)\} \\ &= \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y = Z_\eta y = h; \end{aligned} \tag{8.5}$$

and it can be seen that $y = x + h$. The error dispersion matrices, which are equal, are

$$\begin{aligned} D(\xi|y) &= D(\xi) - C(\xi, y)D^{-1}(y)C(y, \xi) \\ &= \Omega_\xi - \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\xi, \end{aligned} \tag{8.6}$$

$$\begin{aligned} D(\eta|y) &= D(\eta) - C(\eta, y)D^{-1}(y)C(y, \eta), \\ &= \Omega_\eta - \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}\Omega_\eta. \end{aligned} \tag{8.7}$$

The estimating equations can be obtained via the criterion

$$\text{Minimise} \quad S(\xi,\eta)=\xi'\Omega_\xi^{-1}\xi+\eta'\Omega_\eta^{-1}\eta \quad \text{subject to} \quad \xi+\eta=y. \tag{8.8}$$

Since $S(\xi,\eta)$ is the exponent of the normal joint density function, these are the maximum-likelihood estimates.

Setting $\eta=y-\xi$ gives the function

$$S(\xi)=\xi'\Omega_\xi^{-1}\xi+(y-\xi)'\Omega_\eta^{-1}(y-\xi). \tag{8.9}$$

The minimising value is

$$\begin{aligned} x&=(\Omega_\xi^{-1}+\Omega_\eta^{-1})^{-1}\Omega_\eta^{-1}y\\ &=y-\Omega_\eta(\Omega_\eta+\Omega_\xi)^{-1}y=y-h. \end{aligned} \tag{8.10}$$

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

$$(\Omega_\xi+\Omega_\eta)b=y \tag{8.11}$$

for the value of b. Thereafter, one can generate

$$x=\Omega_\xi b \qquad \text{and} \qquad h=\Omega_\eta b. \tag{8.12}$$

If Ω_ξ and Ω_η have a limited number of nonzero bands, then so has the lower-triangular matrix G, which is the Cholesky factor of $\Omega_\xi+\Omega_\eta=GG'$. 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.

Filtering via Fourier Methods

The matrix of the circular autocovariances of the data is obtained by replacing the argument z in the autocovariance generating function $\gamma(z)$ by the matrix $K_T = [e_1, \ldots, e_{T-1}, e_0]$:

$$D^\circ(y) = \Omega^\circ = \gamma(K_T) = \gamma_0 I_T + \sum_{\tau=1}^{\infty} \gamma_\tau (K_T^\tau + K_T^{-\tau}) = \gamma_0^\circ I_T + \sum_{\tau=1}^{T-1} \gamma_\tau^\circ (K_T^\tau + K_T^{-\tau}). \quad (8.13)$$

The circular autocovariances would be obtained by wrapping the sequence of ordinary autocovariances around a circle of circumference T and adding the overlying values. Thus

$$\gamma_\tau^\circ = \sum_{j=0}^{\infty} \gamma_{jT+\tau}, \quad \text{with} \quad \tau = 0, \ldots, T-1. \quad (8.14)$$

The circulant autocovariance matrix is amenable to a spectral factorisation of the form

$$\Omega^\circ = \gamma(K_T) = \bar{U}\gamma(D)U, \quad (8.15)$$

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

$$\gamma(\exp\{\mathrm{i}\omega_j\}) = \gamma_0 + 2\sum_{\tau=1}^{\infty} \gamma_\tau \cos(\omega_j \tau). \quad (8.16)$$

The filtering can also be implemented using the circulant dispersion matrices

$$\begin{aligned}\Omega_\xi^\circ &= \bar{U}\gamma_\xi(D)U, \quad \Omega_\eta^\circ = \bar{U}\gamma_\eta(D)U \quad \text{and}\\ \Omega^\circ &= \Omega_\xi^\circ + \Omega_\eta^\circ = \bar{U}\{\gamma_\xi(D) + \gamma_\eta(D)\}U,\end{aligned} \tag{8.17}$$

where $\gamma_\xi(D)$ and $\gamma_\eta(D)$ contain the ordinates of the component spectra. Replacing the dispersion matrices within (8.5) and (8.6) by their circulant counterparts gives

$$x = \bar{U}\gamma_\xi(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}Uy = P_\xi y, \tag{8.18}$$

$$h = \bar{U}\gamma_\eta(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}Uy = P_\eta y. \tag{8.19}$$

To implement the filters, a Fourier transform is applied to y to give Uy, which is in the frequency domain. Then, the diagonal weighting matrices $J_\xi = \gamma_\xi(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}$ and $J_\eta = \gamma_\eta(D)\{\gamma_\xi(D) + \gamma_\eta(D)\}^{-1}$ are applied. Finally, the products are carried back into the time domain by the inverse Fourier transform, which is represented by the matrix $\bar{U}$.

If the components of $y(t)$ reside in disjoint frequency bands, then these filters can achieve a perfect separation.

Dealing with Trended Data

The trend, estimated via polynomial regression, may be subtracted from the data, or else the data may be reduced to stationarity by taking differences. Both methods entail the matrix difference operator.

The ordinary second-difference operator acting on $x(t) = \{x_t; t = 0, \pm 1, \pm 2, \ldots\}$ is

$$\nabla^2 = (1 - L)^2 = 1 - 2L + L^2, \tag{8.20}$$

where L is such that $Lx(t) = x(t-1)$.

For finite samples, the operator is a submatrix Q' of $\nabla^2(I_T) = \nabla^2_T = [Q_*, Q]'$, illustrated for $T = 6$ as follow:

$$\nabla^2_6 = \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & 0 \\ \hline 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array}\right] = \begin{bmatrix} Q'_* \\ Q' \end{bmatrix}. \tag{8.21}$$

The inverse matrix is $\Sigma_T^2 = [S_*, S]$, and there are

$$[\, S_* \quad S \,] \begin{bmatrix} Q'_* \\ Q' \end{bmatrix} = S_* Q'_* + SQ' = I_T, \tag{8.22}$$

and

$$\begin{bmatrix} Q'_* \\ Q' \end{bmatrix} [\, S_* \quad S \,] = \begin{bmatrix} Q'_* S_* & Q'_* S \\ Q' S_* & Q' S \end{bmatrix} = \begin{bmatrix} I_2 & 0 \\ 0 & I_{T-2} \end{bmatrix}. \tag{8.23}$$

The columns of submatrix S_* provide a basis for the set of all polynomials of degree $d-1$.

For $T = 6$, there is

$$\nabla_6^{-2} = \left[\begin{array}{cc|cccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 & 0 & 0 \\ 4 & 3 & 2 & 1 & 0 & 0 \\ 5 & 4 & 3 & 2 & 1 & 0 \\ 6 & 5 & 4 & 3 & 2 & 1 \end{array}\right] = [\, S_* \quad S \,]. \tag{8.24}$$

The elements of the columns are the coefficients of the binomial expansion of $(1-z)^{-2}$.

The OLS estimator of the trend ordinates is the vector

$$p = S_*(S_*'S_*)^{-1}S_*'y. \tag{8.25}$$

For an alternative expression, we use the identity

$$S_*(S_*'S_*)^{-1}S_*' = I - Q(Q'Q)^{-1}Q', \tag{8.26}$$

which follows since Q and S_* are complementary matrices. This gives

$$p = y - Q(Q'Q)^{-1}Q'y. \tag{8.27}$$

The residual vector is

$$e = Q(Q'Q)^{-1}Q'y. \tag{8.28}$$

Observe that e contains exactly the same information as the vector of differences $Q'y$. The structure of the low-frequency spectrum, which might not be visible in the periodogram of the differences, is evident in the periodogram of the residuals.

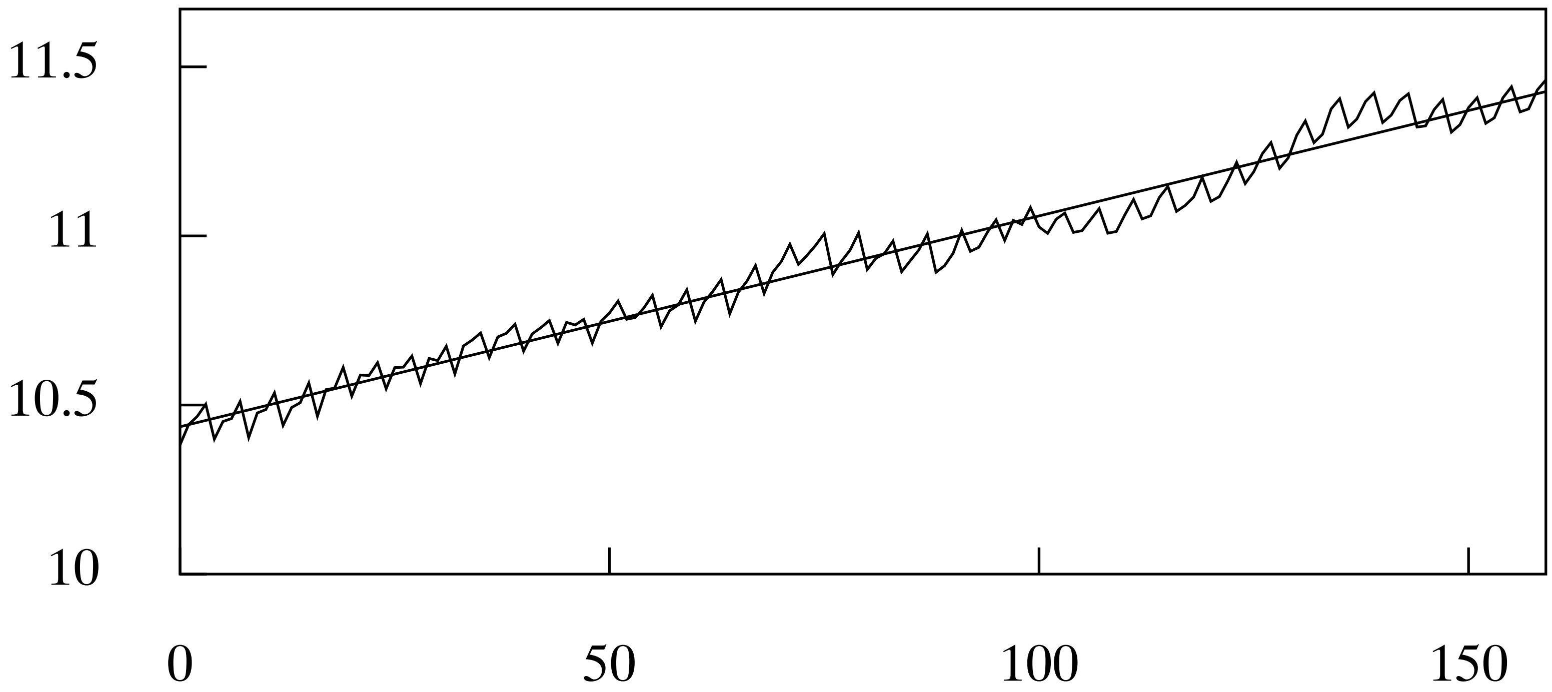

Figure. The quarterly series of the logarithms of consumption in the U.K., for the years 1955 to 1994, together with a quadratic trend interpolated by least-squares regression.

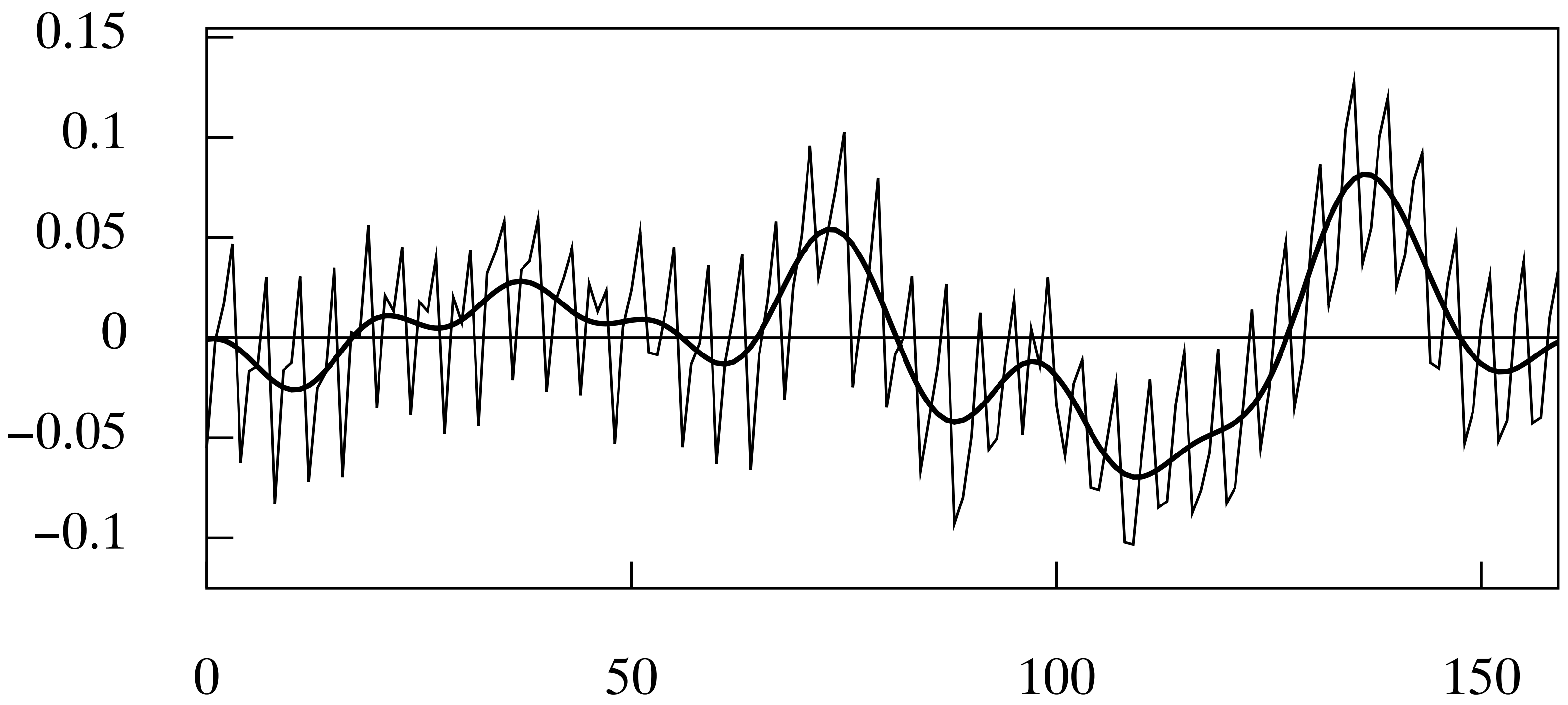

Figure. The residual sequence from fitting a quadratic trend to the logarithmic consumption data. The interpolated line, which represents the business cycle, has been synthesised from the Fourier ordinates in the frequency interval $[0, \pi/8]$.

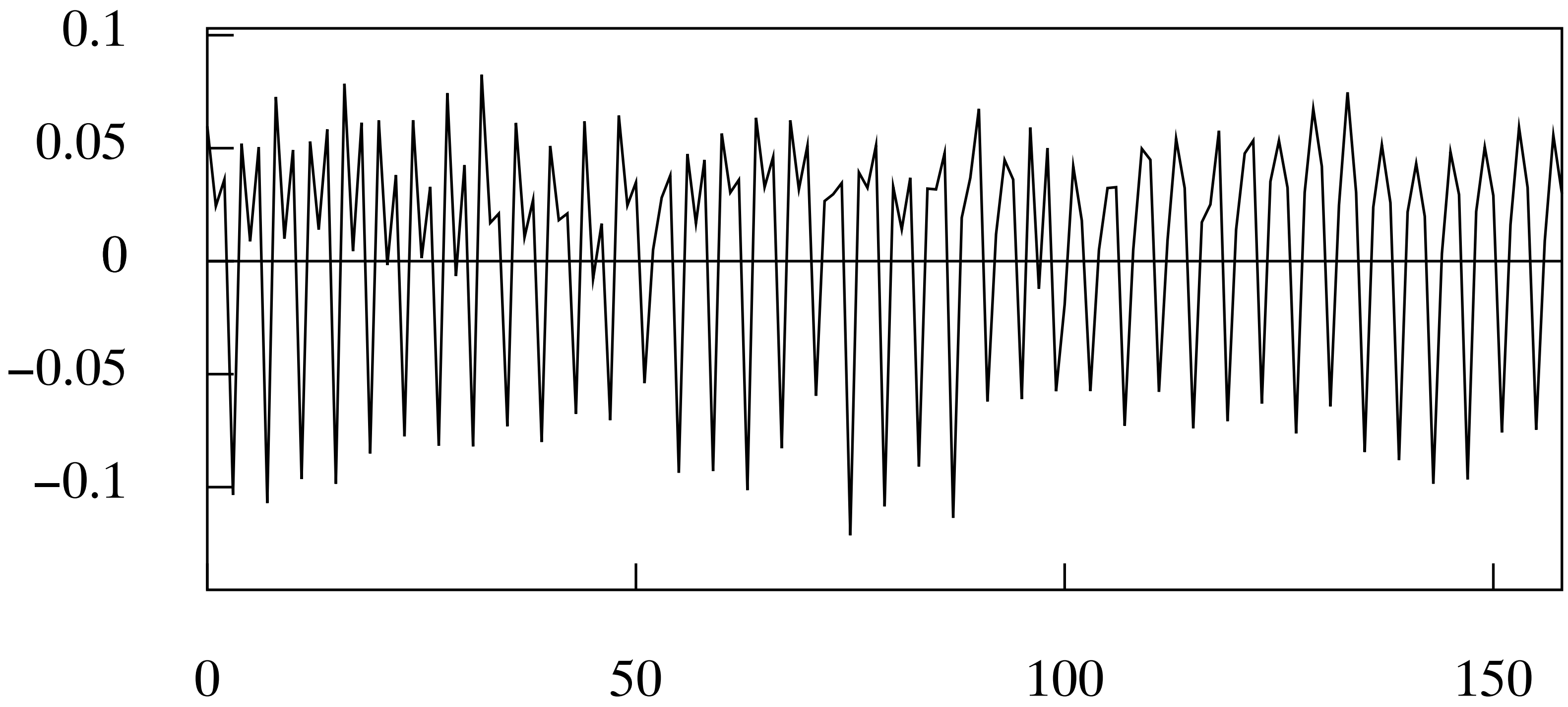

Figure. The differences of the logarithmic consumption data.

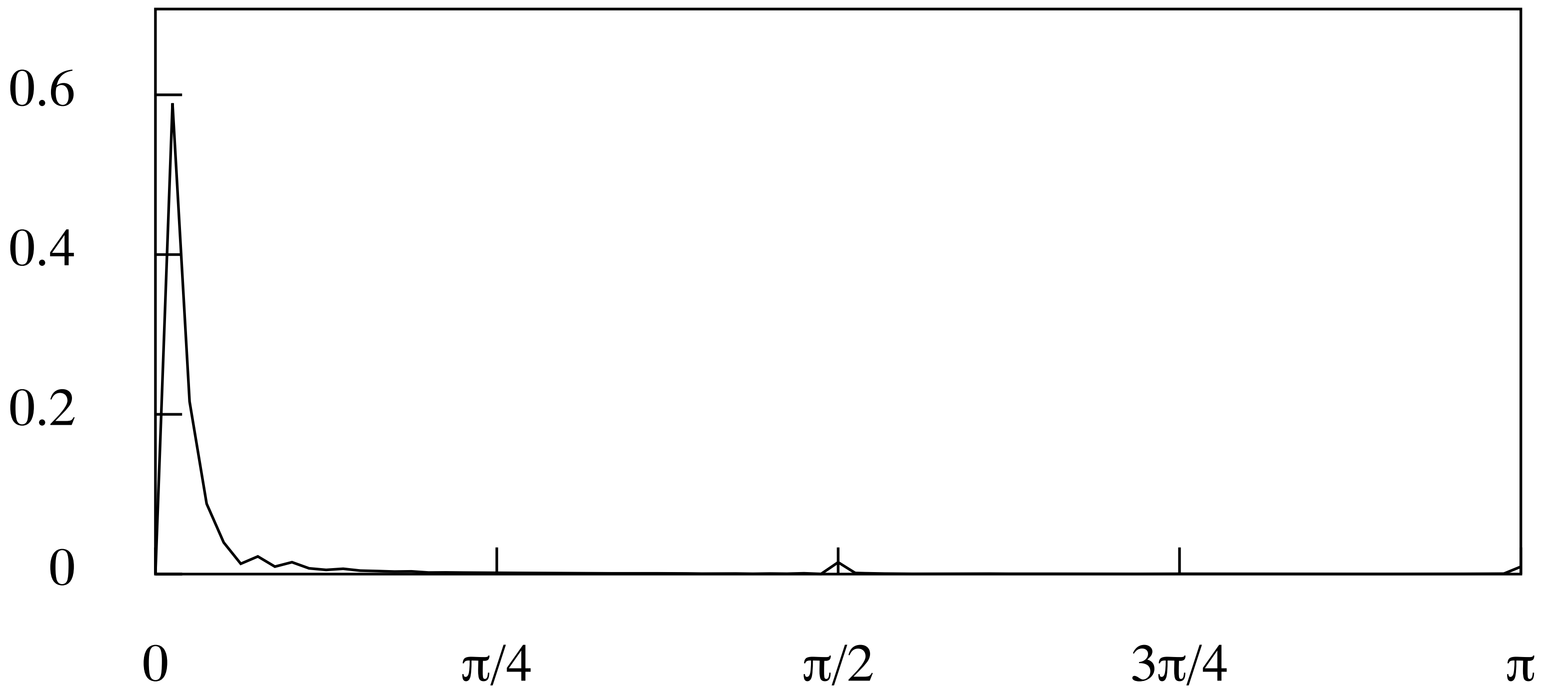

Figure. The periodogram of the logarithms of consumption in the U.K., for the years 1955 to 1994.

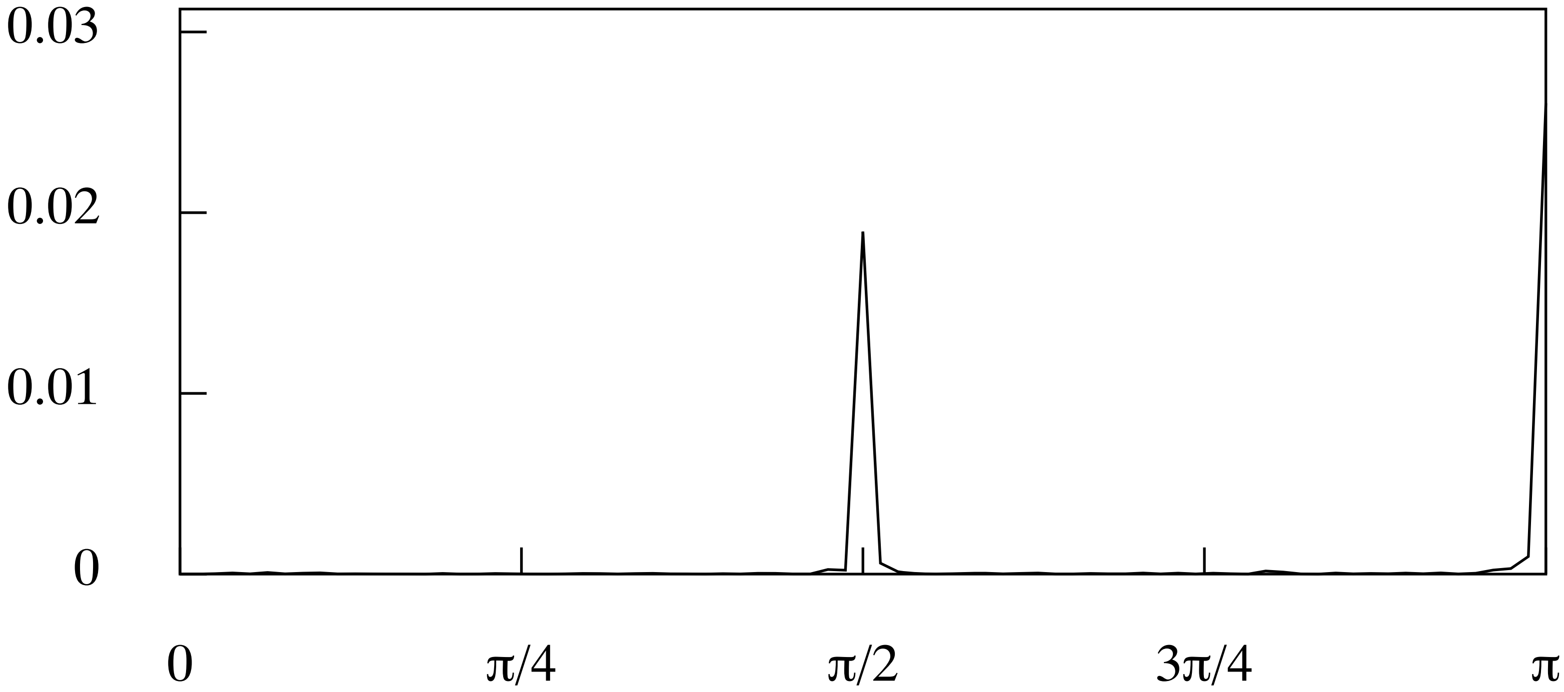

Figure. The periodogram of the first differences of the logarithmic consumption data.

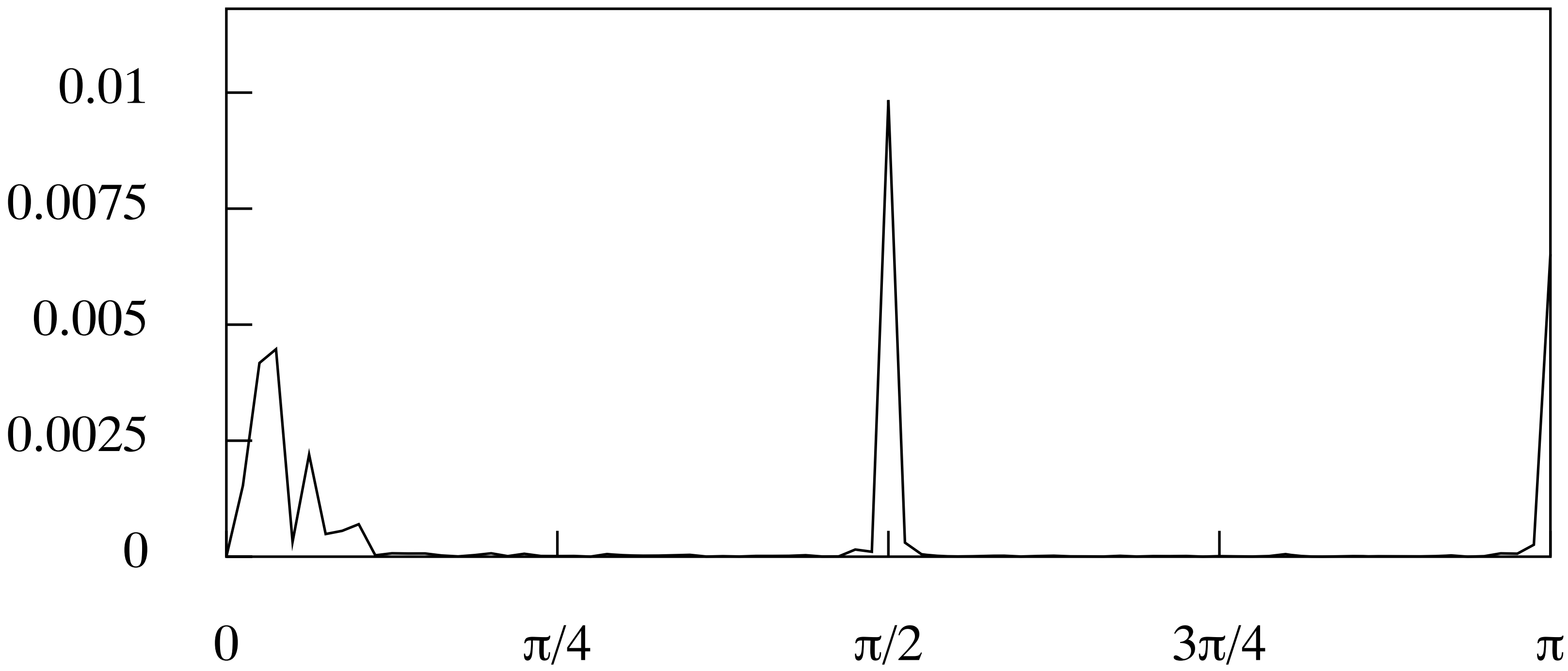

Figure. The periodogram of the residual sequence obtained from the linear detrending of the logarithmic consumption data.

Recovering the Trend Component

The differencing of the data y to give g and the recovery of y from g are represented by

$$\begin{bmatrix} Q'_* \\ Q' \end{bmatrix} y = \begin{bmatrix} g_* \\ g \end{bmatrix} \quad \text{and} \quad y = S_* g_* + Sg, \tag{8.29}$$

where $p = S_* g_*$ is a polynomial of degree $d-1$ over $t = 0, 1, \ldots, T-1$, when d is the degree of differencing. Let

$$Q'y = Q'(x + h) = z + k = g, \tag{8.30}$$

where x is the trend component and h is the residue, and z and k are their differenced versions.

To recover a trend estimate $x = S_* z_* + Sz$ requires the starting values in z_*. The criterion for finding them is

$$\text{Minimise} \quad (y - S_* z_* - Sz)'(y - S_* z_* - Sz) \quad \text{with respect to} \quad z_*. \tag{8.31}$$

The solution is

$$z_* = (S'_* S_*)^{-1} S'_* (y - Sz). \tag{8.32}$$

To recover the residue component $h = Sk + S_* k_*$, the criterion is

$$\text{Minimise} \quad (S_* k_* + Sk)'(S_* k_* + Sk) \quad \text{with respect to} \quad k_*. \tag{8.33}$$

The solution is

$$k_* = -(S'_* S_*)^{-1} S'_* Sk. \tag{8.34}$$

Wiener–Kolmogorov Estimates from Trended Data

The differenced data vector, which is stationary, is

$$Q'y = Q'\xi + Q'\eta = \delta + \kappa = g. \tag{8.35}$$

The dispersion matrices of the differenced components are

$$D(\delta) = \Omega_\delta \quad \text{and} \quad D(\kappa) = Q'\Omega_\eta Q. \tag{8.36}$$

Let the estimates of ξ, η, $\delta = Q'\xi$ and $\kappa = Q'\eta$ be denoted by x, h, d and k respectively. Then

$$d = \Omega_\delta(\Omega_\delta + \Omega_\kappa)^{-1}g = \Omega_\delta(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y, \tag{8.37}$$

$$k = \Omega_\kappa(\Omega_\delta + \Omega_\kappa)^{-1}g = Q'\Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{8.38}$$

It can be shown that the estimate of the residue vector is

$$h = \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y, \tag{8.39}$$

which differs from k by the omission of the leading differencing matrix Q'. The trend estimate x can be obtained from y by the subtraction of the estimated residue:

$$x = y - h = y - \Omega_\eta Q(\Omega_\delta + Q'\Omega_\eta Q)^{-1}Q'y. \tag{8.40}$$

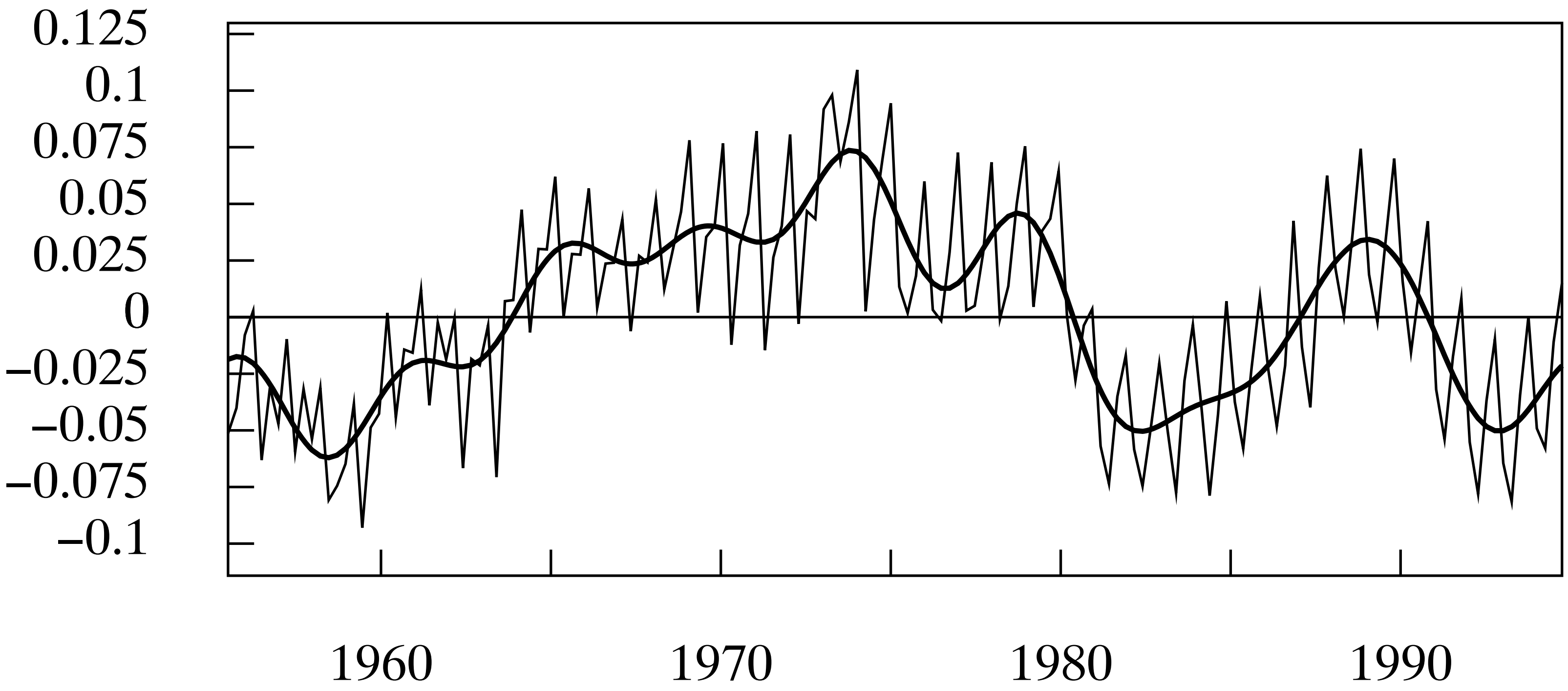

Figure. The residual sequence from fitting a linear trend to the logarithms of the Gross Domestic Product data. The interpolated line, which represents the business cycle, has been synthesised from the Fourier ordinates in the frequency interval $[0, \pi/8]$.

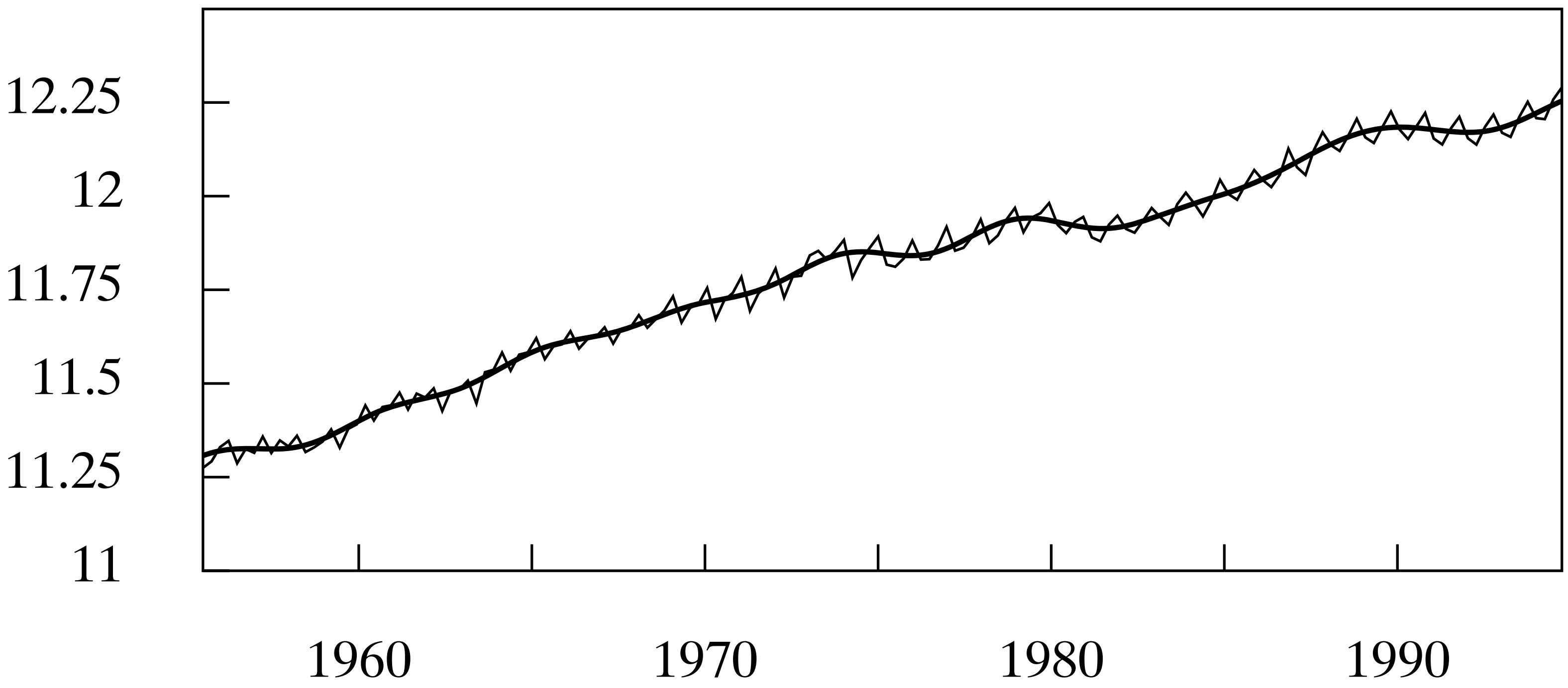

Figure. The trend/cycle component of U.K. Gross Domestic Product determined by the Fourier method, superimposed on the logarithmic data.

A Derivation of the Hodrick–Prescott Filter

It is assumed that the data vector $y = \xi + \eta$, comprises a component ξ generated by a second-order random walk and a component η that is generated by a white-noise process. Recall that Q' denotes the matrix version of the second-difference operator. Then

$$Q'y = Q'\xi + Q'\eta = \zeta + Q'\eta, \tag{8.41}$$

where

$$\begin{aligned} E(\zeta) &= 0, \qquad D(\zeta) = \sigma_\zeta^2 I_{T-2}, \\ E(\eta) &= 0, \qquad D(\eta) = \sigma_\eta^2 I_T, \\ &\text{and} \quad C(\zeta, Q'\eta) = 0. \end{aligned} \tag{8.42}$$

The independence of ξ and η implies that $D(Q'y) = \sigma_\eta^2 Q'Q + \sigma_\zeta^2 I$.

On the assumption that the components have a normal distribution, there is the following joint density function:

$$N(\zeta, \eta) = (2\pi)^{1-T} \sigma_\zeta^{2-T} \sigma_\eta^{-T} \exp\{-\frac{1}{2}(\sigma_\zeta^{-2}\xi'QQ'\xi + \sigma_\eta^{-2}\eta'\eta)\}. \tag{8.43}$$

The maximum-likelihood estimate x of the trend component ξ is found by minimising a criterion function that is derived from the quadratic exponent of the density function by setting $\eta = y - \xi$:

$$S(\xi) = \sigma_\zeta^{-2}\xi' QQ'\xi + \sigma_\eta^{-2}(y-\xi)'(y-\xi). \tag{8.44}$$

The minimising value of ξ is

$$x = \sigma_\eta^{-2}(\sigma_\zeta^{-2}QQ' + \sigma_\eta^{-2}I)^{-1}y. \tag{8.45}$$

According to the matrix inversion lemma, there is

$$(\sigma_\zeta^{-2}QQ' + \sigma_\eta^{-2}I)^{-1} = \sigma_\eta^2\Big\{I - Q(Q'Q + [\sigma_\zeta^2/\sigma_\eta^2]I)\Big\}^{-1}Q'. \tag{8.46}$$

Using this in (8.45) and writing $\sigma_\zeta^2/\sigma_\eta^2 = \lambda^{-1}$, we get

$$x = y - Q(Q'Q + \lambda^{-1}I)^{-1}Q'y. \tag{8.47}$$

This is the appropriate finite-sample version of the H–P trend-estimation filter.