1. EXPECTATIONS AND CONDITIONAL EXPECTATIONS

The joint density function of x and y is

$$f(x, y) = f(x|y)f(y) = f(y|x)f(x), \qquad (1)$$

where

$$f(x) = \int_y f(x, y)dx \qquad \text{and} \qquad f(y) = \int_x f(x, y)dy \qquad (2)$$

are the marginal distributions of x and y respectively and where

$$f(x|y) = \frac{f(y, x)}{f(y)} \qquad \text{and} \qquad f(y|x) = \frac{f(y, x)}{f(x)} \qquad (3)$$

are the conditional distributions of x given y and of y given x.

The unconditional expectation of $y \sim f(y)$ is

$$E(y) = \int_y yf(y)dy. \qquad (4)$$

The conditional expectation of y given x is

$$E(y|x) = \int_y yf(y|x)dy = \int_y y\frac{f(y, x)}{f(x)}dy. \qquad (5)$$

The expectation of the conditional expectation is an unconditional expectation:

$$\begin{aligned}
E\{E(y|x)\} &= \int_x \left\{ \int_y y\frac{f(y, x)}{f(x)}dy \right\} f(x)dx \\
&= \int_x \int_y yf(y, x)f(x)dydx \qquad (6) \\
&= \int_y y\left\{ \int_x f(y, x)dx \right\} dy = \int_y f(y)dy = E(y).
\end{aligned}$$

The conditional expectation of y given x is the minimum mean squared error prediction

Proof. Let $\hat{y} = E(y|x)$ and let $\pi = \pi(x)$ be any other estimator. Then,

$$\begin{aligned}
E\{(y - \pi)^2\} &= E\left[\{(y - \hat{y}) + (\hat{y} - \pi)\}^2 \right] \\
&= E\{(y - \hat{y})^2\} + 2E\{(y - \hat{y})(\hat{y} - \pi)\} + E\{(\hat{y} - \pi)^2\}. \qquad (7)
\end{aligned}$$

In the second term, there is

$$\begin{aligned}
E\{(y - \hat{y})(\hat{y} - \pi)\} &= \int_x \int_y (y - \hat{y})(\hat{y} - \pi)f(x, y)\partial y\partial x \\
&= \int_x \left\{ \int_y (y - \hat{y})f(y|x)\partial y \right\}(\hat{y} - \pi)f(x)\partial x = 0. \qquad (8)
\end{aligned}$$

Therefore, $E\{(y - \pi)^2\} = E\{(y - \hat{y})^2\} + E\{(\hat{y} - \pi)^2\} \geq E\{(y - \hat{y})^2\}$, and the assertion is proved.

The error in predicting y is uncorrelated with x. The proof of this depends on showing that $E(\hat{y}x) = E(yx)$, where $\hat{y} = E(y|x)$:

$$
\begin{aligned}
E(\hat{y}x) &= \int_x x E(y|x) f(x) dx \\
&= \int_x x \left\{ \int_y y \frac{f(y,x)}{f(x)} dy \right\} f(x) dx \\
&= \int_x \int_y xy f(y,x) dy dx = E(xy).
\end{aligned}
\tag{9}
$$

The result can be expressed as $E\{(y - \hat{y})x\} = 0$.

This result can be used in deriving expressions for the parameters α and β of a linear regression of the form

$$
E(y|x) = \alpha + \beta x,
\tag{10}
$$

from which and unconditional expectation is derived in the form of

$$
E(y) = \alpha + \beta E(x).
\tag{11}
$$

The orthogonality of the prediction error implies that

$$
\begin{aligned}
0 = E\{(y - \hat{y})x\} &= E\{(y - \alpha - \beta x)x\} \\
&= E(xy) - \alpha E(x) - \beta E(x^2).
\end{aligned}
\tag{12}
$$

In order to eliminate $\alpha E(x)$ from this expression, equation (11) is multiplied by $E(x)$ and rearranged to give

$$
\alpha E(x) = E(x)E(y) - \beta \{E(x)\}^2.
\tag{13}
$$

This substituted into (12) to give

$$
E(xy) - E(x)E(y) = \beta \big[E(x^2) - \{E(x)\}^2 \big],
\tag{14}
$$

whence

$$
\beta = \frac{E(xy) - E(x)E(y)}{E(x^2) - \{E(x)\}^2} = \frac{C(x,y)}{V(x)}.
\tag{15}
$$

The expression

$$
\alpha = E(y) - \beta E(x)
\tag{16}
$$

comes directly from (9).

Observe that, by substituting (16) into (10), the following prediction-error equation for the conditional expectation is derived:

$$
E(y|x) = E(y) + \beta \{x - E(x)\}.
\tag{17}
$$

2

Thus, the conditional expectation of y given x is obtained by adjusting the unconditional expectation by some proportion of the error in predicting x by its expected value.

2. THE PARTITIONED REGRESSSION MODEL

Consider taking a regression equation in the form of

$$y = [\,X_1 \quad X_2\,] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{1}$$

Here $[X_1, X_2] = X$ and $[\beta_1', \beta_2']' = \beta$ are obtained by partitioning the matrix X and vector β of the equation $y = X\beta + \varepsilon$ in a conformable manner. The normal equations $X'X\beta = X'y$ can be partitioned likewise. Writing the equations without the surrounding matrix braces gives

$$X_1'X_1\beta_1 + X_1'X_2\beta_2 = X_1'y, \tag{2}$$

$$X_2'X_1\beta_1 + X_2'X_2\beta_2 = X_2'y. \tag{3}$$

From (2), we get the equation $X_1'X_1\beta_1 = X_1'(y - X_2\beta_2)$ which gives an expression for the leading subvector of $\hat{\beta}$:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2). \tag{4}$$

To obtain an expression for $\hat{\beta}_2$, we must eliminate β_1 from equation (3). For this purpose, we multiply equation (2) by $X_2'X_1(X_1'X_1)^{-1}$ to give

$$X_2'X_1\beta_1 + X_2'X_1(X_1'X_1)^{-1}X_1'X_2\beta_2 = X_2'X_1(X_1'X_1)^{-1}X_1'y. \tag{5}$$

When the latter is taken from equation (3), we get

$$\left\{X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\right\}\beta_2 = X_2'y - X_2'X_1(X_1'X_1)^{-1}X_1'y. \tag{6}$$

On defining

$$P_1 = X_1(X_1'X_1)^{-1}X_1', \tag{7}$$

can we rewrite (6) as

$$\left\{X_2'(I - P_1)X_2\right\}\beta_2 = X_2'(I - P_1)y, \tag{8}$$

whence

$$\hat{\beta}_2 = \left\{X_2'(I - P_1)X_2\right\}^{-1}X_2'(I - P_1)y. \tag{9}$$

Now let us investigate the effect that conditions of orthogonality amongst the regressors have upon the ordinary least-squares estimates of the regression parameters. Consider a partitioned regression model, which can be written as

$$y = [\,X_1, X_2\,] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon. \tag{10}$$

It can be assumed that the variables in this equation are in deviation form. Imagine that the columns of X_1 are orthogonal to the columns of X_2 such that $X_1'X_2 = 0$. This is the same as assuming that the empirical correlation between variables in X_1 and variables in X_2 is zero.

The effect upon the ordinary least-squares estimator can be seen by examining the partitioned form of the formula $\hat{\beta} = (X'X)^{-1}X'y$. Here we have

$$X'X = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \begin{bmatrix} X_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}, \qquad (11)$$

where the final equality follows from the condition of orthogonality. The inverse of the partitioned form of $X'X$ in the case of $X_1'X_2 = 0$ is

$$(X'X)^{-1} = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix}. \qquad (12)$$

We also have

$$X'y = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}. \qquad (13)$$

On combining these elements, we find that

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1}X_1'y \\ (X_2'X_2)^{-1}X_2'y \end{bmatrix}. \qquad (14)$$

In this special case, the coefficients of the regression of y on $X = [X_1, X_2]$ can be obtained from the separate regressions of y on X_1 and y on X_2.

It should be understood that this result does not hold true in general. The general formulae for $\hat{\beta}_1$ and $\hat{\beta}_2$ are those which we have given already under (4) and (9):

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2),$$
$$\hat{\beta}_2 = \left\{ X_2'(I - P_1)X_2 \right\}^{-1} X_2'(I - P_1)y, \quad P_1 = X_1(X_1'X_1)^{-1}X_1'. \qquad (15)$$

It can be confirmed easily that these formulae do specialise to those under (14) in the case of $X_1'X_2 = 0$.

The purpose of including X_2 in the regression equation when, in fact, interest is confined to the parameters of β_1 is to avoid falsely attributing the explanatory power of the variables of X_2 to those of X_1.

Let us investigate the effects of erroneously excluding X_2 from the regression. In that case, the estimate will be

$$\begin{aligned} \tilde{\beta}_1 &= (X_1'X_1)^{-1}X_1'y \\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon) \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon. \end{aligned} \qquad (16)$$

On applying the expectations operator to these equations, we find that

$$E(\tilde{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2, \tag{17}$$

since $E\{(X_1'X_1)^{-1}X_1'\varepsilon\} = (X_1'X_1)^{-1}X_1'E(\varepsilon) = 0$. Thus, in general, we have $E(\tilde{\beta}_1) \neq \beta_1$, which is to say that $\tilde{\beta}_1$ is a biased estimator. The only circumstances in which the estimator will be unbiased are when either $X_1'X_2 = 0$ or $\beta_2 = 0$. In other circumstances, the estimator will suffer from a problem which is commonly described as *omitted-variables bias.*

We need to ask whether it matters that the estimated regression parameters are biased. The answer depends upon the use to which we wish to put the estimated regression equation. The issue is whether the equation is to be used simply for predicting the values of the dependent variable y or whether it is to be used for some kind of structural analysis.

If the regression equation purports to describe a structural or a behavioral relationship within the economy, and if some of the explanatory variables on the RHS are destined to become the instruments of an economic policy, then it is important to have unbiased estimators of the associated parameters. For these parameters indicate the leverage of the policy instruments. Examples of such instruments are provided by interest rates, tax rates, exchange rates and the like.

On the other hand, if the estimated regression equation is to be viewed solely as a predictive device—that it to say, if it is simply an estimate of the function $E(y|x_1, \ldots, x_k)$ which specifies the conditional expectation of y given the values of $x_1, \ldots, x_n$—then, provided that the underlying statistical mechanism which has generated these variables is preserved, the question of the unbiasedness the regression estimates does not arise.

3. HYPOTHESES CONCERNING SUBSETS OF THE REGRESSION COEFFICIENTS

Consider a set of linear restrictions on the vector β of a classical linear regression model $N(y; X\beta, \sigma^2 I)$ which take the form of

$$R\beta = r, \tag{1}$$

where R is a matrix of order $j \times k$ and of rank j, which is to say that the j restrictions are independent of each other and are fewer in number than the parameters within β. We know that the ordinary least-squares estimator of β is a normally distributed vector $\hat{\beta} \sim N\{\beta, \sigma^2(X'X)^{-1}\}$. It follow that

$$R\hat{\beta} \sim N\{R\beta = r, \sigma^2 R(X'X)^{-1}R'\}; \tag{2}$$

and, from this, we can immediately infer that

$$\frac{(R\hat{\beta} - r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r)}{\sigma^2} \sim \chi^2(j). \tag{3}$$

We have already established the result that

$$\frac{(T-k)\hat{\sigma}^2}{\sigma^2} = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{\sigma^2} \sim \chi^2(T-k) \tag{4}$$

is a chi-square variate which is statistically independent of the chi-square variate

$$\frac{(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)}{\sigma^2} \sim \chi^2(k) \tag{5}$$

derived from the estimator of the regression parameters. The variate of (4) must also be independent of the chi-square of (3); and it is straightforward to deduce that

$$F = \left\{ \frac{(R\hat{\beta} - r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r)}{j} \middle/ \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T-k} \right\}$$

$$= \frac{(R\hat{\beta} - r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r)}{\hat{\sigma}^2 j} \sim F(j, T-k), \tag{6}$$

which is to say that the ratio of the two independent chi-square variates, divided by their respective degrees of freedom, is an F statistic. This statistic, which embodies only know and observable quantities, can be used in testing the validity of the hypothesised restrictions $R\beta = r$.

A specialisation of the statistic under (6) can also be used in testing an hypothesis concerning a subset of the elements of the vector β. Let $\beta' = [\beta_1', \beta_2']'$. Then the condition that the subvector β_1 assumes the value of β_1^* can be expressed via the equation

$$[I_{k_1}, 0] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \beta_1^*. \tag{7}$$

This can be construed as a case of the equation $R\beta = r$ where $R = [I_{k_1}, 0]$ and $r = \beta_1^*$.

In order to discover the specialised form of the requisite test statistic, let us consider the following partitioned form of an inverse matrix:

$$(X'X)^{-1} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \{X_1'(I - P_2)X_1\}^{-1} & -\{X_1'(I - P_2)X_1\}^{-1}X_1'X_2(X_2'X_2)^{-1} \\ -\{X_2'(I - P_1)X_2\}^{-1}X_2'X_1(X_1'X_1)^{-1} & \{X_2'(I - P_1)X_2\}^{-1} \end{bmatrix}, \tag{8}$$

Then, with $R = [I, 0]$, we find that

$$R(X'X)^{-1}R' = \{X_1'(I - P_2)X_1\}^{-1} \tag{9}$$

It follows in a straightforward manner that the specialised form of the F statistic of (6) is

$$
\begin{aligned}
F &= \left\{ \frac{(\hat{\beta}_1 - \beta_1^*)'\{X_1'(I - P_2)X_1\}(\hat{\beta}_1 - \beta_1^*)}{k_1} \middle/ \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T - k} \right\} \\
&= \frac{(\hat{\beta}_1 - \beta_1^*)'\{X_1'(I - P_2)X_1'\}(\hat{\beta}_1 - \beta_1^*)}{\hat{\sigma}^2 k_1} \sim F(k_1, T - k).
\end{aligned}
\tag{10}
$$

By specialising the expression under (10), a statistic may be derived for testing the hypothesis that $\beta_i = \beta_{i\diamond}$, concerning a single element:

$$
F = \frac{(\hat{\beta}_i - \beta_{i\diamond})^2}{\hat{\sigma}^2 w_{ii}},
\tag{11}
$$

Here, w_{ii} stands for the ith diagonal element of $(X'X)^{-1}$. If the hypothesis is true, then this will have an $F(1, T - k)$ distribution.

However, the usual way of testing such an hypothesis is to use

$$
t = \frac{\hat{\beta}_i - \beta_{i\diamond}}{\sqrt{(\hat{\sigma}^2 w_{ii})}}
\tag{12}
$$

in conjunction with the tables of the $t(T-k)$ distribution. The t statistic shows the direction in which the estimate of β_i deviates from the hypothesised value as well as the size of the deviation.

4. REGRESSIONS ON TRIGONOMETRICAL FUNCTIONS: DISCRETE-TIME FOURIER ANALYSIS

An example of orthogonal regressors is a Fourier analysis, where the explanatory variables are sampled from a set of trigonometric functions with angular velocities, called Fourier frequencies, that are evenly distributed in an interval from zero to π radians per sample period.

If the sample is indexed by $t = 0, 1, \ldots, T-1$, then the Fourier frequencies are $\omega_j = 2\pi j/T$; $j = 0, 1, \ldots, [T/2]$, where $[T/2]$ denotes the integer quotient of the division of T by 2.

The object of a Fourier analysis is to express the elements of the sample as a weighted sum of sine and cosine functions as follows:

$$
y_t = \alpha_0 + \sum_{j=1}^{[T/2]} \{\alpha_j \cos(\omega_j t) + \beta \sin(\omega_j t)\}; \quad t = 0, 1, \ldots, T-1.
\tag{1}
$$

The vectors of the generic trigonometric regressors may be denoted by

$$
c_j = [c_{0j}, c_{1j}, \ldots c_{T-1,j}]' \quad \text{and} \quad s_j = [s_{0j}, s_{1j}, \ldots s_{T-1,j}]',
\tag{3}
$$

7

where $c_{tj} = \cos(\omega_j t)$ and $s_{tj} = \sin(\omega_j t)$. The vectors of the ordinates of functions of different frequencies are mutually orthogonal. Therefore, the following orthogonality conditions hold:

$$c_i' c_j = s_i' s_j = 0 \quad \text{if} \quad i \neq j,$$
$$\text{and} \quad c_i' s_j = 0 \quad \text{for all} \quad i, j. \tag{4}$$

In addition, there are some sums of squares which can be taken into account in computing the coefficients of the Fourier decomposition:

$$c_0' c_0 = \iota' \iota = T, \qquad s_0' s_0 = 0,$$
$$c_j' c_j = s_j' s_j = T/2 \quad \text{for} \quad j = 1, \ldots, [(T-1)/2] \tag{5}$$

When $T = 2n$, there is $\omega_n = \pi$ and there is also

$$s_n' s_n = 0, \qquad \text{and} \qquad c_n' c_n = T. \tag{6}$$

The "regression" formulae for the Fourier coefficients can now be given. First, there is

$$\alpha_0 = (\iota' \iota)^{-1} \iota' y = \frac{1}{T} \sum_t y_t = \bar{y}. \tag{7}$$

Then, for $j = 1, \ldots, [(T-1)/2]$, there are

$$\alpha_j = (c_j' c_j)^{-1} c_j' y = \frac{2}{T} \sum_t y_t \cos \omega_j t, \tag{8}$$

and

$$\beta_j = (s_j' s_j)^{-1} s_j' y = \frac{2}{T} \sum_t y_t \sin \omega_j t. \tag{9}$$

If $T = 2n$ is even, then there is no coefficient β_n and there is

$$\alpha_n = (c_n' c_n)^{-1} c_n' y = \frac{1}{T} \sum_t (-1)^t y_t. \tag{10}$$

By pursuing the analogy of multiple regression, it can be seen, in view of the orthogonality relationships, that there is a complete decomposition of the sum of squares of the elements of the vector y:

$$y' y = \alpha_0^2 \iota' \iota + \sum_{j=1}^{[T/2]} \left\{ \alpha_j^2 c_j' c_j + \beta_j^2 s_j' s_j \right\}. \tag{11}$$

Now consider writing $\alpha_0^2 \iota' \iota = \bar{y}^2 \iota' \iota = \bar{y}' \bar{y}$, where $\bar{y}' = [\bar{y}, \bar{y}, \ldots, \bar{y}]$ is a vector whose repeated element is the sample mean $\bar{y}$. It follows that $y' y - \alpha_0^2 \iota' \iota =$

$y'y - \bar{y}'\bar{y} = (y - \bar{y})'(y - \bar{y})$. Then, in the case where $T = 2n$ is even, the equation can be written as

$$(y - \bar{y})'(y - \bar{y}) = \frac{T}{2} \sum_{j=1}^{n-1} \left\{\alpha_j^2 + \beta_j^2\right\} + T\alpha_n^2 = \frac{T}{2} \sum_{j=1}^{n} \rho_j^2. \qquad (12)$$

where $\rho_j = \alpha_j^2 + \beta_j^2$ for $j = 1, \ldots, n-1$ and $\rho_n = 2\alpha_n$. A similar expression exists when T is odd, with the exceptions that α_n is missing and that the summation runs to $(T-1)/2$.

It follows that the variance of the sample can be expressed as

$$\frac{1}{T} \sum_{t=0}^{T-1} (y_t - \bar{y})^2 = \frac{1}{2} \sum_{j=1}^{n} (\alpha_j^2 + \beta_j^2). \qquad (13)$$

The proportion of the variance that is attributable to the component at frequency ω_j is $(\alpha_j^2 + \beta_j^2)/2 = \rho_j^2/2$, where ρ_j is the amplitude of the component.

The number of the Fourier frequencies increases at the same rate as the sample size T, and, if there are no regular harmonic components in the underling process, then we can expect the proportion of the variance attributed to the individual frequencies to decline as the sample size increases.

If there is a regular component, then we can expect the the variance attributable to it to converge to a finite value as the sample size increases.

In order provide a graphical representation of the decomposition of the sample variance, we must scale the elements of equation (36) by a factor of T. The graph of the function $I(\omega_j) = (T/2)(\alpha_j^2 + \beta_j^2)$ is know as the periodogram.

We should note that the trigonometric functions that underlie the periodogram are defined over an infinite index set $\{t = 0 \pm 1, \pm 2, \ldots\}$ comprising all positive and negative integers. As the index t progresses through this sequence, equation (1) will generate successive replications of the data set that is defined on the points $t = 0, 1, \ldots, T-1$. The result is described as the periodic extension of the data.

Since the trigonometric functions are both periodic and strictly bounded, they cannot be used to synthesise a perpetual trend. Therefore, before a periodogram analysis can be applied effectively, it is necessary to ensure that the data sequence is free of trend. A failure to detrend the data will lead to successive disjunctions in its periodic extension at the points where the beginning of one replication of the data set joins the end of the previous replication. The result will be a so-called saw tooth function.

The typical periodogram of a saw tooth function has the form of a rectangular hyperbola, which descends from a high point at the fundamental frequency of $\omega_1 = 2\pi/T$ and which reaches a low point in the vicinity of the limiting frequency of π. Within such a periodogram, the features of primary interest, which relate to the fluctuations that are superimposed upon the trend, may be virtually imperceptible—see Figure 5, for example.
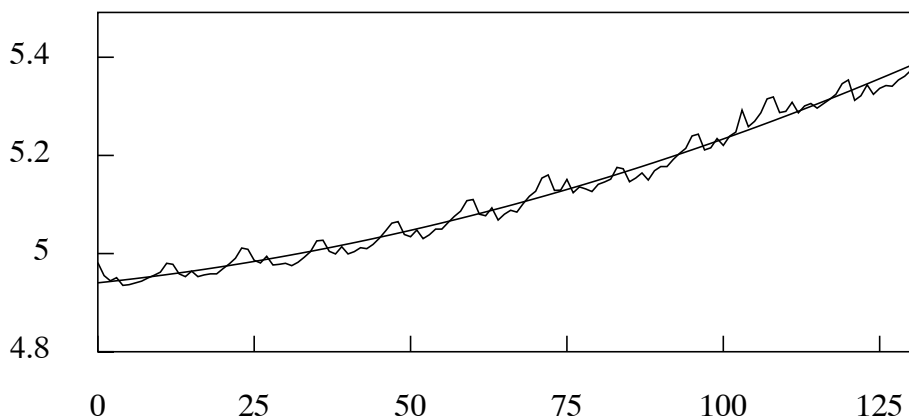
9

Figure 1. The plot of 132 monthly observations on the U.S. money supply, beginning in January 1960. A quadratic function has been interpolated through the data.
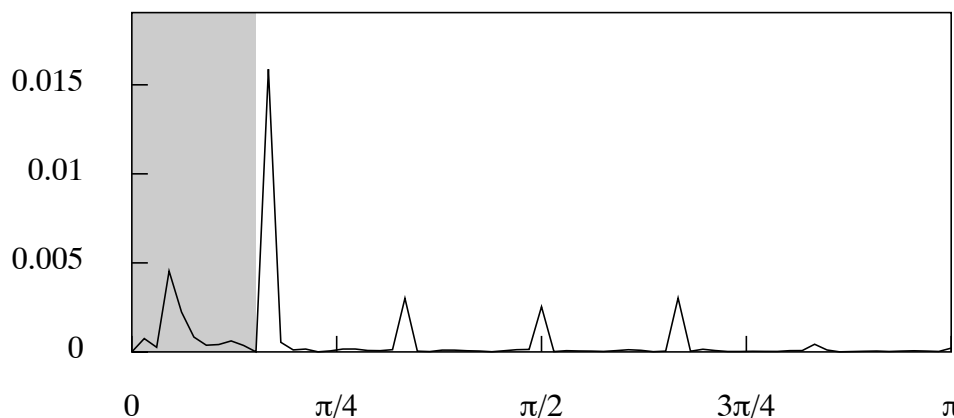

Figure 2. The periodogram of the residuals of the logarithmic money-supply data.

The Periodogram and the Autocovariance Function

A stationary stochastic process can be characterised, equivalently, by its autocovariance function or its partial autocovariance function.

It can also be characterised by is spectral density function, which is the Fourier transform of the autocovariances $\{\gamma_\tau; \tau = 0, \pm 1, \pm 2, \ldots\}$:

$$f(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau \cos(\omega\tau) = \gamma_0 + 2\sum_{\tau=1}^{\infty} \gamma_\tau \cos(\omega\tau). \qquad (14)$$

Here, $\omega \in [0, \pi]$ is an angular velocity, or frequency value, in radians per period.

The empirical counterpart of the spectral density function is the periodogram $I(\omega_j)$, which may be defined as

$$\frac{1}{2}I(\omega_j) = \sum_{\tau=1-T}^{T-1} c_\tau \cos(\omega_j\tau) = c_0 + 2\sum_{\tau=1}^{T-1} c_\tau \cos(\omega_j\tau), \qquad (15)$$

10

where $\{c_\tau; \tau = 0, \pm 1, \ldots, \pm(T-1)\}$, with

$$c_\tau = T^{-1} \sum_{t=\tau}^{T-1} (y_t - \bar{y})(y_{t-\tau} - \bar{y}), \tag{16}$$

are the empirical autocovariances, and where $\omega_j = 2\pi j/T; j = 0, 1, \ldots, [T/2]$ are the Fourier frequencies and We need to show this definition of the peridogram is equivalent to the previous definition, which was based on the following frequency decomposition of the sample variance:

$$\frac{1}{T} \sum_{t=0}^{T-1} (y_t - \bar{y})^2 = \frac{1}{2} \sum_{j=0}^{[T/2]} (\alpha_j^2 + \beta_j^2), \tag{17}$$

where

$$\alpha_j = \frac{2}{T} \sum_t y_t \cos(\omega_j t) = \frac{2}{T} \sum_t (y_t - \bar{y}) \cos(\omega_j t),$$

$$\beta_j = \frac{2}{T} \sum_t y_t \sin(\omega_j t) = \frac{2}{T} \sum_t (y_t - \bar{y}) \sin(\omega_j t).$$

Substituting these into the term $T(\alpha_j^2 + \beta_j^2)/2$ gives the periodogram

$$I(\omega_j) = \frac{2}{T} \left[\left\{ \sum_{t=0}^{T-1} \cos(\omega_j t)(y_t - \bar{y}) \right\}^2 + \left\{ \sum_{t=0}^{T-1} \sin(\omega_j t)(y_t - \bar{y}) \right\}^2 \right].$$

The quadratic terms may be expanded to give

$$I(\omega_j) = \frac{2}{T} \left\{ \sum_t \sum_s \cos(\omega_j t) \cos(\omega_j s)(y_t - \bar{y})(y_s - \bar{y}) \right\}$$
$$+ \frac{2}{T} \left\{ \sum_t \sum_s \sin(\omega_j t) \sin(\omega_j s)(y_t - \bar{y})(y_s - \bar{y}) \right\},$$

Since $\cos(A)\cos(B) + \sin(A)\sin(B) = \cos(A - B)$, this can be written as

$$I(\omega_j) = \frac{2}{T} \left\{ \sum_t \sum_s \cos(\omega_j[t - s])(y_t - \bar{y})(y_s - \bar{y}) \right\}$$

On defining $\tau = t - s$ and writing $c_\tau = \sum_t (y_t - \bar{y})(y_{t-\tau} - \bar{y})/T$, we can reduce the latter expression to

$$I(\omega_j) = 2 \sum_{\tau=1-T}^{T-1} \cos(\omega_j \tau) c_\tau,$$

which is a Fourier transform of the empirical autocovariances.

5. FILTERING MACROECONOMIC 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,

$$E(\xi) = 0, \qquad D(\xi) = \Omega_\xi,$$

$$E(\eta) = 0, \qquad D(\eta) = \Omega_\eta, \tag{2}$$

$$\text{and} \quad C(\xi, \eta) = 0.$$

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:

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

This indicates that the estimate 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.
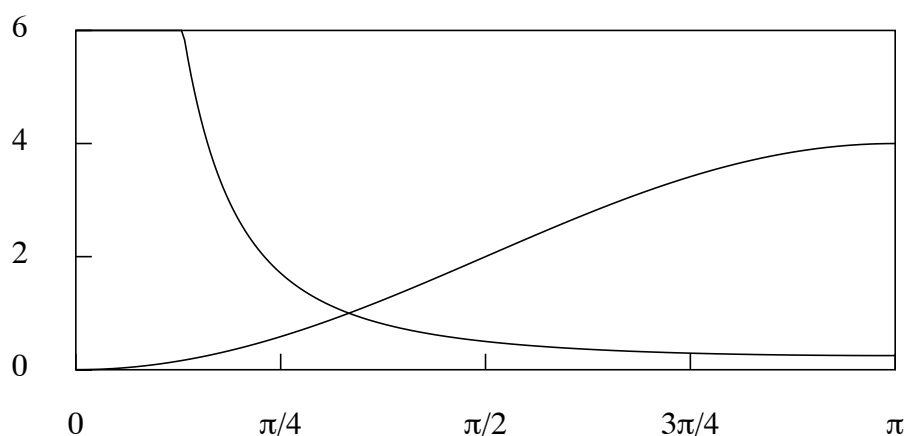
Figure 3. The squared gain of the difference operator, which has a zero at zero frequency, and the squared gain of the summation operator, which is unbounded at zero frequency.

The Matrix Difference Operator

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

$$\nabla_5^2 = \begin{bmatrix} Q'_* \\ Q' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix}. \tag{9}$$

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
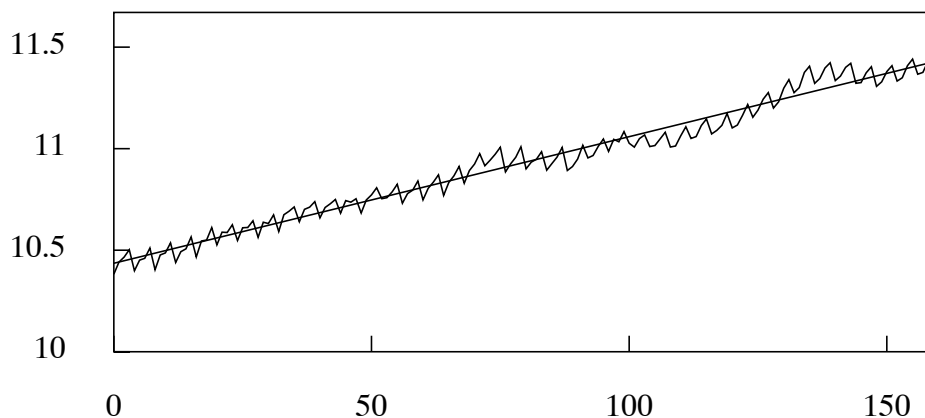
14

Figure 4. 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.
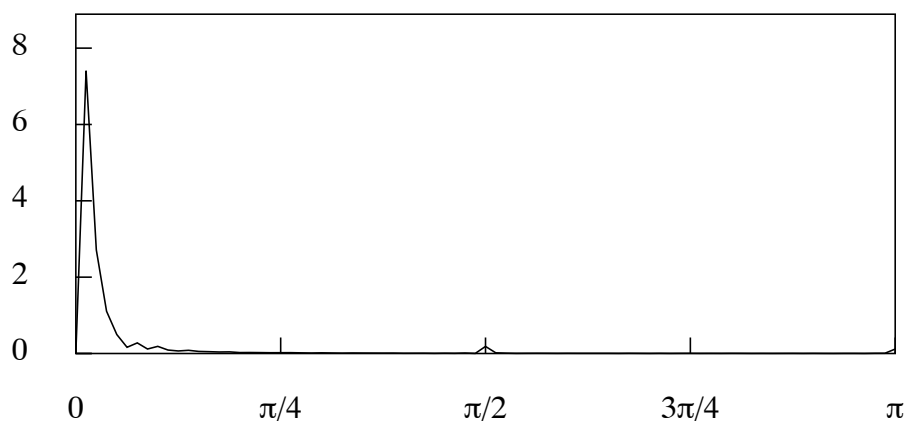

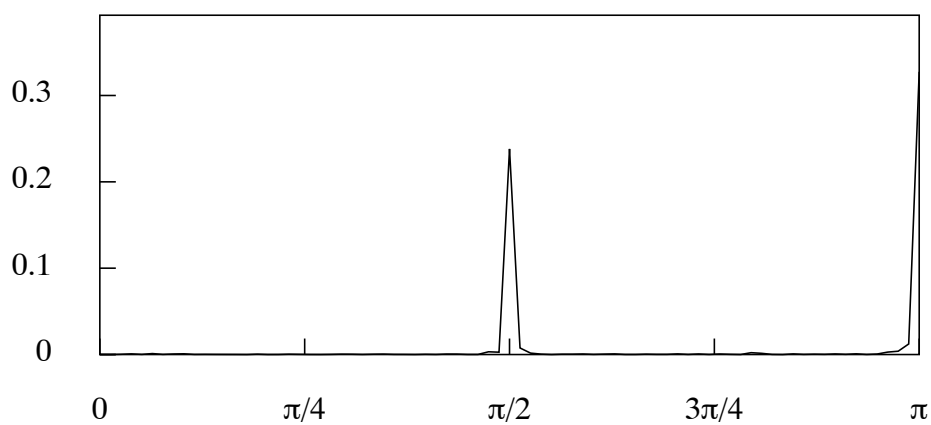**Figure 5.** The periodogram of the trended logarithmic data.

Figure 6. The periodogram of the differenced logarithmic consumpption data.
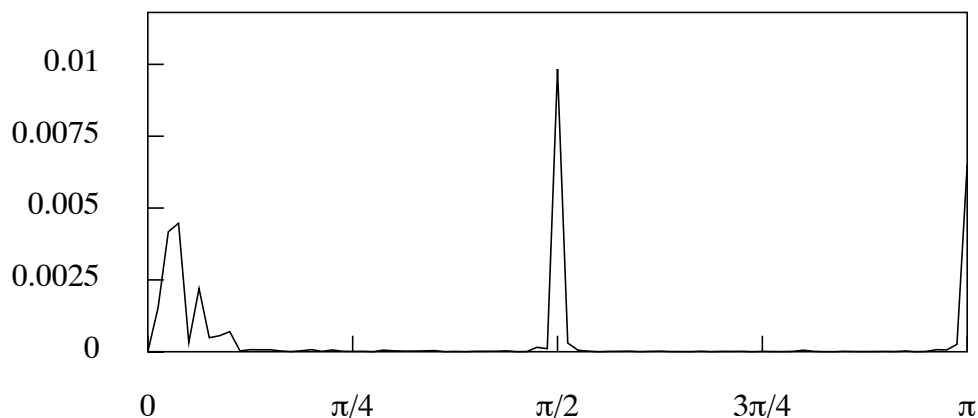
Figure 7. The periodogram of the residual sequence obtained from the linear detrending of the logarithmic consumption data.

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

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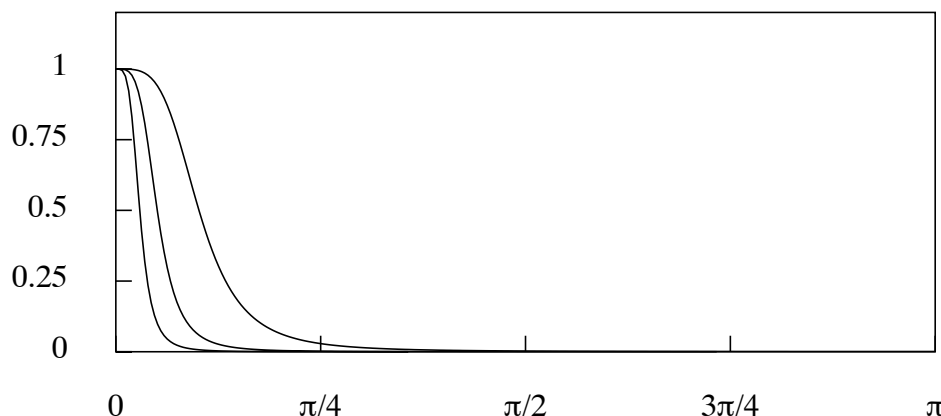$$

16

Figure 8. 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):
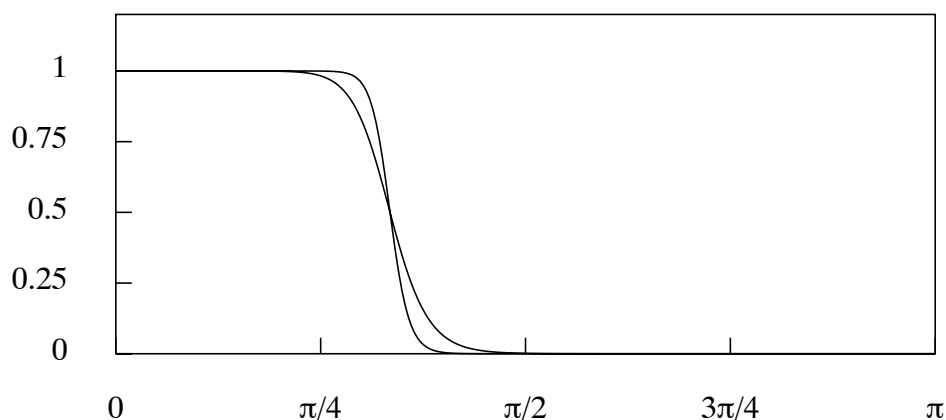
$$x = y - Q(Q'Q)^{-1} Q' y.$$

Figure 9. 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 8 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 9 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.

LAGGED DEPENDENT VARIABLES AND
ERROR CORRECTION MECHANISMS

A regression equation can also be set in motion by including lagged values of the dependent variable on the RHS. With one lagged value, we get

$$y(t) = \phi y(t-1) + \beta x(t) + \varepsilon(t). \tag{1}$$

In terms of the lag operator, this is

$$(1 - \phi L)y(t) = \beta x(t) + \varepsilon(t), \tag{2}$$

of which the rational form is

$$y(t) = \frac{\beta}{1 - \phi L}x(t) + \frac{1}{1 - \phi L}\varepsilon(t). \tag{3}$$

The advantage of equation (1) is that it is amenable to estimation by ordinary least-squares regression. Although the estimates will be biased in finite samples, they will be consistent, if the model is correctly specified.

The disadvantage is the restrictive assumption that the systematic and disturbance parts have the same dynamics.

Partial Adjustment and Adaptive Expectations

A simple partial-adjustment model has the form

$$y(t) = \lambda\{\gamma x(t)\} + (1 - \lambda)y(t-1) + \varepsilon(t), \tag{4}$$

If $y(t)$ is current consumption, $x(t)$ is disposable income, then $\gamma x(t) = y^*(t)$ is "desired"consumption. If habits of consumption persist, then current consumption will be a weighted combination of the previous consumption and present desired consumption.

The weights of the combination depend on the partial-adjustment parameter $\lambda \in (0, 1]$. If $\lambda = 1$, then the consumers adjust their consumption instantaneously to the desired value. As $\lambda \to 0$, their consumption habits become increasingly persistent.

When the notation $\lambda\gamma = \beta$ and $(1 - \lambda) = \phi$ is adopted, equation (4) becomes identical to equation (2), which is the regression model with a lagged dependent variable. Observe also that $\gamma = \beta/(1-\phi)$ is a "long-term multipier" in the relationhip between iincome and comsumption. It can also be described as the steady-state gain of the transfer function from $x(t)$ to $y(t)$, which is depocted in equation (3).

Error-Correction Forms, and Nonstationary Signals

The usual linear regression procedures presuppose that the relevant data moments will converge asymptotically to fixed limits as the sample size increases. This cannot happen if the data are trended, in which case, the standard techniques of statistical inference will not be applicable.

A common approach is to subject the data to as many differencing operations as may be required to achieve stationarity. However, differencing tends to remove some of the essential information regarding the behaviour of economic agents. Moreover, it is often discovered that the regression model looses much of its explanatory power when the differences of the data are used instead.

In such circumstances, one might use the so-called error-correction model. The model depicts a mechanism whereby two trended economic variables maintain an enduring long-term proportionality with each other.

The data sequences comprised by the model are stationary, either individually or in an appropriate combination; and this enables us apply the standard procedures of statistical inference that are appropriate to models comprising data from stationary processes.

Consider taking $y(t-1)$ from both sides of the equation of (1), which represents the first-order dynamic model. This gives

$$
\begin{aligned}
\nabla y(t) = y(t) - y(t-1) &= (\phi - 1)y(t-1) + \beta x(t) + \varepsilon(t) \\
&= (1 - \phi)\left\{ \frac{\beta}{1 - \phi} x(t) - y(t-1) \right\} + \varepsilon(t) \qquad (5) \\
&= \lambda\big\{ \gamma x(t) - y(t-1) \big\} + \varepsilon(t),
\end{aligned}
$$

where $\lambda = 1 - \phi$ and where γ is the gain of the transfer function from $x(t)$ to $y(t)$ defined under (3). This is the so-called error-correction form of the equation; and it indicates that the change in $y(t)$ is a function of the extent to which the proportions of the series $x(t)$ and $y(t-1)$ differs from those which would prevail in the steady state.

The error-correction form provides the basis for estimating the parameters of the model when the signal series $x(t)$ is trended or nonstationary. A pair of nonstationary series that maintain a long-run proportionality are said to be cointegrated. It is easy to obtain an accurate estimate of γ, which is the coefficient of proportionality, simply by running a regression of $y(t-1)$ on $x(t)$.

Once a value for γ is available, the remaining parameter λ may be estimated by regressing $\nabla y(t)$ upon the composite variable $\{\gamma x(t) - y(t-1)\}$. However, if the error-correction model is an unrestricted reparametrisation of an original model in levels, then its parameters can be estimated by ordinary least-squares regression. The same estimates can also be inferred from the least-squares estimates of the parameters of the original model in levels. A pair of nonstationary series that maintain a long-run proportionality are said to be cointegrated. It is easy to obtain an accurate estimate of γ, which is the coefficient of proportionality, simply by running a regression of $y(t-1)$ on $x(t)$.

To see how to derive an error-correction form for a more general autoregressive distributed-lag model, consider the second-order model:

$$
y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \beta_0 x(t) + \beta_1 x(t-1) + \varepsilon(t). \qquad (6)
$$

The part $\phi_1 y(t-1) + \phi_2 y(t-2)$ comprising the lagged dependent variables can be reparameterised as follows:

$$\left\{ \begin{bmatrix} \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right\} \left\{ \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y(t-1) \\ y(t-2)) \end{bmatrix} \right\} = \begin{bmatrix} \theta & \rho \end{bmatrix} \begin{bmatrix} y(t-1) \\ \nabla y(t-1)) \end{bmatrix}. \quad (7)$$

Here, the matrix that postmultiplies the row vector of the parameters is the inverse of the matrix that premultiplies the column vector of the variables.

The sum $\beta_0 x(t) + \beta_1 x(t-1)$ can be reparametrised to become

$$\left\{ \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \right\} \left\{ \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-1) \end{bmatrix} \right\} = \begin{bmatrix} \kappa & \delta \end{bmatrix} \begin{bmatrix} x(t-1) \\ \nabla x(t) \end{bmatrix}. \quad (8)$$

It follows that equation (61) can be recast in the form of

$$y(t) = \theta y(t-1) + \rho \nabla y(t-1) + \kappa x(t-1) + \delta \nabla x(t) + \varepsilon(t). \quad (9)$$

Taking $y(t-1)$ from both sides of this equation and rearranging it gives

$$\begin{aligned} \nabla y(t) &= (1-\theta) \left\{ \frac{\kappa}{1-\theta} x(t-1) - y(t-1) \right\} + \rho \nabla y(t-1) + \delta \nabla x(t) + \varepsilon(t) \\ &= \lambda \left\{ \gamma x(t-1) - y(t-1) \right\} + \rho \nabla y(t-1) + \delta \nabla x(t) + \varepsilon(t). \end{aligned} \quad (10)$$

This is an elaboration of equation (5); and it includes the differenced sequences $\nabla y(t-1)$ and $\nabla x(t)$. These are deemed to be stationary, as is the composite error sequence $\gamma x(t-1) - y(t-1)$.

Observe that, in contrast to equation (5), the error-correction term of (10) comprises the lagged value $x(t-1)$ in place of $x(t)$. Had the reparametrising transformation that has been employed in equation (7) also been used in (8), then the consequence would have been to generate an error-correction term of the form $\gamma x(t) - y(t-1)$. It should also be observed that the parameter associated with $x(t)$ in (10), which is

$$\gamma = \frac{\kappa}{1-\phi} = \frac{\beta_0 + \beta_1}{1 - \phi_1 - \phi_2}, \quad (11)$$

is the steady state gain of the transfer function from $x(t)$ to $y(t)$.

Additional lagged differences can be added to the equation (10); and this is tantamount to increasing the number of lags of the dependent variable $y(t)$ and the number of lags of the input variable $x(t)$ within equation (6).