THE MULTIPLE REGRESSION MODEL

Consider T realisations of the regression equation

$$(1) \qquad y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon,$$

which can be written in the following form:

$$(2) \qquad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{T1} & \ldots & x_{Tk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}.$$

This can be represented in summary notation by

$$(3) \qquad y = X\beta + \varepsilon.$$

The object is to derive an expression for the ordinary least-squares estimates of the elements of the parameter vector $\beta = [\beta_0, \beta_1, \ldots, \beta_k]'$.

The ordinary least-squares (OLS) estimate of β is the value that minimises

$$\begin{aligned} S(\beta) &= \varepsilon'\varepsilon \\ &= (y - X\beta)'(y - X\beta) \\ &= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta \\ &= y'y - 2y'X\beta + \beta'X'X\beta. \end{aligned} \tag{4}$$

According to the rules of matrix differentiation, the derivative is

$$\frac{\partial S}{\partial \beta} = -2y'X + 2\beta'X'X. \tag{5}$$

Setting this to zero gives $0 = \beta'X'X - y'X$, which is transposed to provide the so-called normal equations:

$$X'X\beta = X'y. \tag{6}$$

On the assumption that the inverse matrix exists, the equations have a unique solution, which is the vector of ordinary least-squares estimates:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{7}$$

The Decomposition of the Sum of Squares

The equation $y = X\hat{\beta} + e$, decomposes y into a regression component $X\hat{\beta}$ and a residual component $e = y - \hat{X}\beta$. These are mutually orthogonal, since (6) indicates that $X'(y - \hat{X}\beta) = 0$.

Define the projection matrix $P = X(X'X)^{-1}X'$, which is symmetric and idempotent such that

$$P = P' = P^2 \quad \text{or, equivalently,} \quad P'(I - P) = 0.$$

Then, $X\hat{\beta} = Py$ and $e = y - \hat{X}\beta = (I - P)y$, and, therefore, the regression decomposition is

$$y = Py + (I - P)y.$$

The conditions on P imply that

$$\begin{aligned} y'y &= \{Py + (I - P)y\}'\{Py + (I - P)y\} \\ &= y'Py + y'(I - P)y \\ &= \hat{\beta}'X'X\hat{\beta} + e'e. \end{aligned} \tag{8}$$

This is an instance of Pythagorus theorem; and the equation indicates that the total sum of squares $y'y$ is equal to the regression sum of squares $\hat{\beta}'X'X\hat{\beta}$ plus the residual or error sum of squares $e'e$.

By projecting y perpendicularly onto the manifold of X, the distance between y and $Py = X\hat{\beta}$ is minimised.

Proof. Let $\gamma = Pg$ be an arbitrary vector in the manifold of X. Then

$$(9)\qquad \begin{aligned}(y-\gamma)'(y-\gamma) &= \big\{(y - X\hat{\beta}) + (X\hat{\beta} - \gamma)\big\}'\big\{(y - X\hat{\beta}) + (X\hat{\beta} - \gamma)\big\}\\ &= \big\{(I-P)y + P(y-g)\big\}'\big\{(I-P)y + P(y-g)\big\}.\end{aligned}$$

The properties of P indicate that

$$(10)\qquad \begin{aligned}(y-\gamma)'(y-\gamma) &= y'(I-P)y + (y-g)'P(y-g)\\ &= e'e + (X\hat{\beta} - \gamma)'(X\hat{\beta} - \gamma).\end{aligned}$$

Since the squared distance $(X\hat{\beta} - \gamma)'(X\hat{\beta} - \gamma)$ is nonnegative, it follows that $(y-\gamma)'(y-\gamma) \geq e'e$, where $e = y - X\hat{\beta}$; which proves the assertion.

The Coefficient of Determination

A summary measure of the extent to which the ordinary least-squares regression accounts for the observed vector y is provided by the coefficient of determination. This is defined by

$$R^2 = \frac{\hat{\beta}' X' X \hat{\beta}}{y'y} = \frac{y' P y}{y'y}. \tag{11}$$

The measure is just the square of the cosine of the angle between the vectors y and $Py = X\hat{\beta}$; and the inequality $0 \leq R^2 \leq 1$ follows from the fact that the cosine of any angle must lie between -1 and $+1$.

If X is a square matrix of full rank, with as many regressors as observations, then X^{-1} exists and

$$P = X(X'X)^{-1}X = X\{X^{-1}X'^{-1}\}X' = I,$$

and so $R^2 = 1$. If $X'y = 0$, then, $Py = 0$ and $R^2 = 0$. But, if y is distibuted continuously, then this event has a zero probability.

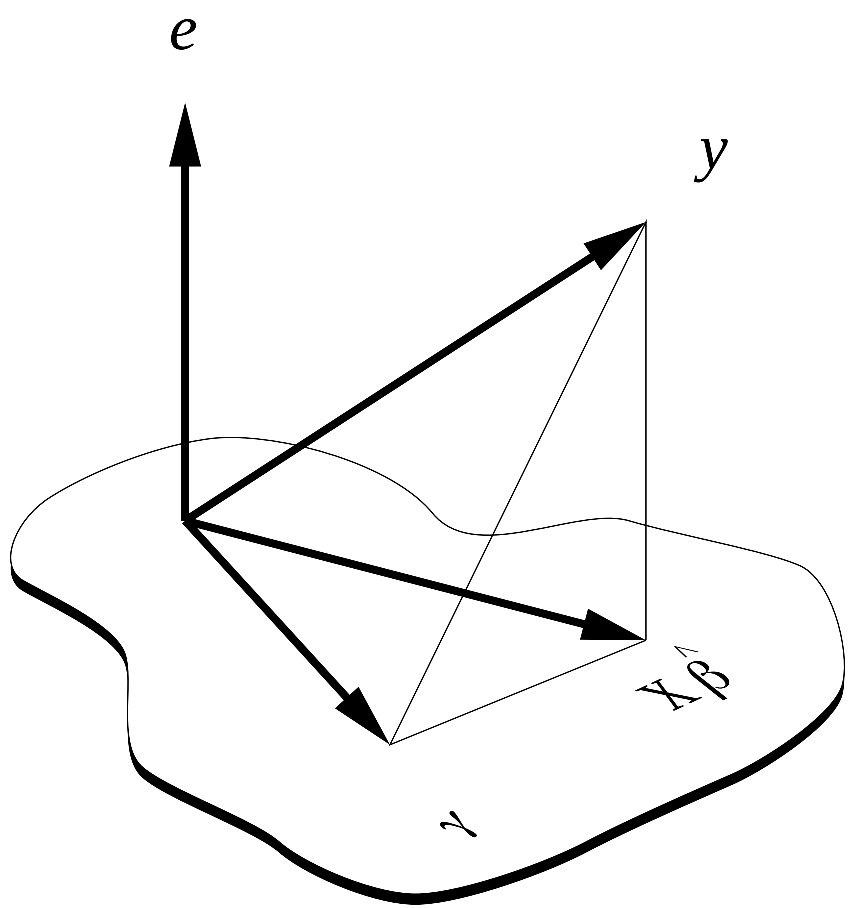

Figure 1. The vector $Py = X\hat{\beta}$ is formed by the orthogonal projection of the vector y onto the subspace spanned by the columns of the matrix X.

The Partitioned Regression Model

Consider partitioning the regression equation of (3) to give

$$(12)\qquad y = [\,X_1 \quad X_2\,]\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon,$$

where $[X_1, X_2] = X$ and $[\beta_1', \beta_2']' = \beta$. The normal equations of (6) can be partitioned likewise:

$$(13)\qquad X_1'X_1\beta_1 + X_1'X_2\beta_2 = X_1'y,$$

$$(14)\qquad X_2'X_1\beta_1 + X_2'X_2\beta_2 = X_2'y.$$

From (13), we get the $X_1'X_1\beta_1 = X_1'(y - X_2\beta_2)$, which gives

$$(15)\qquad \hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2).$$

To obtain an expression for $\hat{\beta}_2$, we must eliminate β_1 from equation (14). For this, we multiply equation (13) by $X_2'X_1(X_1'X_1)^{-1}$ to give

$$(16)\qquad X_2'X_1\beta_1 + X_2'X_1(X_1'X_1)^{-1}X_1'X_2\beta_2 = X_2'X_1(X_1'X_1)^{-1}X_1'y.$$

From

$$X_2'X_1\beta_1 + X_2'X_2\beta_2 = X_2'y, \tag{14}$$

we take the resulting equation

$$X_2'X_1\beta_1 + X_2'X_1(X_1'X_1)^{-1}X_1'X_2\beta_2 = X_2'X_1(X_1'X_1)^{-1}X_1'y \tag{16}$$

to give

$$\Big\{X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\Big\}\beta_2 = X_2'y - X_2'X_1(X_1'X_1)^{-1}X_1'y. \tag{17}$$

On defining $P_1 = X_1(X_1'X_1)^{-1}X_1'$, equation (17) can be written as

$$\Big\{X_2'(I - P_1)X_2\Big\}\beta_2 = X_2'(I - P_1)y, \tag{19}$$

whence

$$\hat{\beta}_2 = \Big\{X_2'(I - P_1)X_2\Big\}^{-1}X_2'(I - P_1)y. \tag{20}$$

The Regression Model with an Intercept

Consider again the equations

$$y = \iota\alpha + Z\beta_z + \varepsilon. \tag{22}$$

where $\iota = [1, 1, \ldots, 1]'$ is the summation vector and $Z = [x_{tj}]$, with $t = 1, \ldots T$ and $j = 1, \ldots, k$, is the matrix of the explanatory variables.

This is a case of the partitioned regression equation of (12). By setting $X_1 = \iota$ and $X_2 = Z$ and by taking $\beta_1 = \alpha$, $\beta_2 = \beta_z$, the equations (15) and (20), give the following estimates of the α and β_z:

$$\hat{\alpha} = (\iota'\iota)^{-1}\iota'(y - Z\hat{\beta}_z), \tag{23}$$

and

$$\begin{aligned}\hat{\beta}_z =& \big\{Z'(I - P_\iota)Z\big\}^{-1}Z'(I - P_\iota)y, \quad \text{with} \\ P_\iota =& \iota(\iota'\iota)^{-1}\iota' = \frac{1}{T}\iota\iota'.\end{aligned} \tag{24}$$

To understand the effect of the operator P_ι, consider

$$(25)\qquad \iota' y = \sum_{t=1}^{T} y_t, \qquad (\iota'\iota)^{-1}\iota' y = \frac{1}{T}\sum_{t=1}^{T} y_t = \bar{y},$$

$$\text{and}\quad P_\iota y = \iota\bar{y} = \iota(\iota'\iota)^{-1}\iota' y = [\bar{y}, \bar{y}, \ldots, \bar{y}]'.$$

Here, $P_\iota y = [\bar{y}, \bar{y}, \ldots, \bar{y}]'$ is a column vector containing T repetitions of the sample mean.

From the above, it can be understood that, if $x = [x_1, x_2, \ldots x_T]'$ is vector of T elements, then

$$(26)\qquad x'(I - P_\iota)x = \sum_{t=1}^{T} x_t(x_t - \bar{x}) = \sum_{t=1}^{T}(x_t - \bar{x})x_t = \sum_{t=1}^{T}(x_t - \bar{x})^2.$$

The final equality depends on the fact that $\sum(x_t - \bar{x})\bar{x} = \bar{x}\sum(x_t - \bar{x}) = 0$.

The Regression Model in Deviation Form

Consider the matrix of cross-products in equation (24). This is

$$Z'(I - P_\iota)Z = \{(I - P_\iota)Z\}'\{Z(I - P_\iota)\} = (Z - \bar{Z})'(Z - \bar{Z}). \tag{27}$$

Here, $\bar{Z}$ contains the sample means of the k explanatory variables repeated T times. The matrix $(I - P_\iota)Z = (Z - \bar{Z})$ contains the deviations of the data points about the sample means. The vector $(I - P_\iota)y = (y - \iota\bar{y})$ may be described likewise.

It follows that the estimate $\hat{\beta}_z = \left\{Z'(I - P_\iota)Z\right\}^{-1} Z'(I - P_\iota)y$ is obtained by applying the least-squares regression to the equation

$$\begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_T - \bar{y} \end{bmatrix} = \begin{bmatrix} x_{11} - \bar{x}_1 & \dots & x_{1k} - \bar{x}_k \\ x_{21} - \bar{x}_1 & \dots & x_{2k} - \bar{x}_k \\ \vdots & & \vdots \\ x_{T1} - \bar{x}_1 & \dots & x_{Tk} - \bar{x}_k \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 - \bar{\varepsilon} \\ \varepsilon_2 - \bar{\varepsilon} \\ \vdots \\ \varepsilon_T - \bar{\varepsilon} \end{bmatrix}, \tag{28}$$

which lacks an intercept term.

In summary notation, the equation may be denoted by

$$y - \iota\bar{y} = [Z - \bar{Z}]\beta_z + (\varepsilon - \bar{\varepsilon}). \tag{29}$$

Observe that it is unnecessary to take the deviations of y. The result is the same whether we regress y or $y - \iota\bar{y}$ on $[Z - \bar{Z}]$. The result is due to the symmetry and idempotency of the operator $(I - P_\iota)$, whereby $Z'(I - P_\iota)y = \{(I - P_\iota)Z\}'\{(I - P_\iota)y\}$.

Once the value for $\hat{\beta}_z$ is available, the estimate for the intercept term can be recovered from the equation (23), which can be written as

$$\hat{\alpha} = \bar{y} - \bar{Z}\hat{\beta}_z = \bar{y} - \sum_{j=1}^{k} \bar{x}_j\hat{\beta}_j. \tag{30}$$

The Assumptions of the Classical Linear Model

Consider the regression equation

$$y = X\beta + \varepsilon, \tag{32}$$

where $y = [y_1, y_2, \ldots, y_T]'$, $\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T]'$, $\beta = [\beta_0, \beta_1, \ldots, \beta_k]'$ and $X = [x_{tj}]$, with $x_{t0} = 1$ for all t.

It is assumed that the disturbances have expected values of zero. Thus

$$E(\varepsilon) = 0 \qquad \text{or, equivalently,} \qquad E(\varepsilon_t) = 0, \quad t = 1, \ldots, T. \tag{33}$$

Next, it is assumed that they are mutually uncorrelated and that they have a common variance. Thus

$$D(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I, \qquad \text{or} \qquad E(\varepsilon_t\varepsilon_s) = \begin{cases} \sigma^2, & \text{if } t = s; \\ 0, & \text{if } t \neq s. \end{cases} \tag{34}$$

If t is a temporal index, then these assumptions imply that there is no inter-temporal correlation in the sequence of disturbances.

A conventional assumption, borrowed from the experimental sciences, is that X is a nonstochastic matrix with linearly independent columns.

Linear independence is necessary in order to distinguish the separate effects of the k explanatory variables.

In econometrics, it is more appropriate to regard the elements of X as random variables distributed independently of the disturbances:

$$E(X'\varepsilon|X) = X'E(\varepsilon) = 0. \tag{37}$$

Then,

$$\hat{\beta} = (X'X)^{-1}X'y \quad \text{is unbiased such that} \quad E(\hat{\beta}) = \beta. \tag{38}$$

To demonstrate this, we may write

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= \beta + (X'X)^{-1}X'\varepsilon.\end{aligned} \tag{39}$$

Taking expectations gives

$$\begin{aligned}E(\hat{\beta}) &= \beta + (X'X)^{-1}X'E(\varepsilon) \\ &= \beta.\end{aligned} \tag{40}$$

Notice that, in the light of this result, equation (39) now indicates that

$$\hat{\beta} - E(\hat{\beta}) = (X'X)^{-1}X'\varepsilon. \tag{41}$$

The variance–covariance matrix of the ordinary least-squares regression estimator is

$$D(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

This is demonstrated via the following sequence of identities:

$$\begin{aligned} E\Big\{\big[\hat{\beta} - E(\hat{\beta})\big]\big[\hat{\beta} - E(\hat{\beta})\big]'\Big\} &= E\{(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\} \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\ &= (X'X)^{-1}X'\{\sigma^2 I\}X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}. \end{aligned} \tag{43}$$

The second of these equalities follows directly from equation (41).

Matrix Traces

If $A = [a_{ij}]$ is a square matrix, then $\text{Trace}(A) = \sum_{i=1}^{n} a_{ii}$. If $A = [a_{ij}]$ is of order $n \times m$ and $B = [b_{k\ell}]$ is of order $m \times n$, then

$$(45)\qquad \begin{aligned} AB = C = [c_{i\ell}] \quad &\text{with} \quad c_{i\ell} = \sum_{j=1}^{m} a_{ij} b_{j\ell} \quad \text{and} \\ BA = D = [d_{kj}] \quad &\text{with} \quad d_{kj} = \sum_{\ell=1}^{n} b_{k\ell} a_{\ell j}. \end{aligned}$$

Now,

$$(46)\qquad \begin{aligned} \text{Trace}(AB) &= \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ji} \quad \text{and} \\ \text{Trace}(BA) &= \sum_{j=1}^{m} \sum_{\ell=1}^{n} b_{j\ell} a_{\ell j} = \sum_{\ell=1}^{n} \sum_{j=1}^{m} a_{\ell j} b_{j\ell}. \end{aligned}$$

Apart from a change of notation, where ℓ replaces i, the expressions on the RHS are the same. It follows that $\text{Trace}(AB) = \text{Trace}(BA)$. For three factors A, B, C, we have $\text{Trace}(ABC) = \text{Trace}(CAB) = \text{Trace}(BCA)$.

Estimating the Variance of the Disturbance

It is natural to estimate $\sigma^2 = V(\varepsilon_t)$ via its empirical counterpart. With $e_t = y_t - x_{t.}\hat{\beta}$ in place of ε_t, it follows that $T^{-1}\sum_t e_t^2$ may be used to estimate σ^2. However, it transpires that this is biased. An unbiased estimate is provided by

$$(48) \qquad \hat{\sigma}^2 = \frac{1}{T-k}\sum_{t=1}^{T} e_t^2 = \frac{1}{T-k}(y - X\hat{\beta})'(y - X\hat{\beta}).$$

The unbiasedness of this estimate may be demonstrated by finding the expected value of $(y - X\hat{\beta})'(y - X\hat{\beta}) = y'(I - P)y$.

Given that $(I - P)y = (I - P)(X\beta + \varepsilon) = (I - P)\varepsilon$ in consequence of the condition $(I - P)X = 0$, it follows that

$$(49) \qquad E\Big\{(y - X\hat{\beta})'(y - X\hat{\beta})\Big\} = E(\varepsilon'\varepsilon) - E(\varepsilon' P\varepsilon).$$

The value of the first term on the RHS is given by

$$E(\varepsilon'\varepsilon) = \sum_{t=1}^{T} E(e_t^2) = T\sigma^2. \tag{50}$$

The value of the second term on the RHS is given by

$$\begin{aligned} E(\varepsilon' P\varepsilon) &= \text{Trace}\big\{E(\varepsilon' P\varepsilon)\big\} = E\big\{\text{Trace}(\varepsilon' P\varepsilon)\big\} = E\big\{\text{Trace}(\varepsilon\varepsilon' P)\big\} \\ &= \text{Trace}\big\{E(\varepsilon\varepsilon')P\big\} = \text{Trace}\big\{\sigma^2 P\big\} = \sigma^2\text{Trace}(P) \\ &= \sigma^2 k. \end{aligned} \tag{51}$$

The final equality follows from the fact that $\text{Trace}(P) = \text{Trace}(I_k) = k$. Putting the results of (50) and (51) into (49), gives

$$E\big\{(y - X\hat{\beta})'(y - X\hat{\beta})\big\} = \sigma^2(T - k); \tag{52}$$

and, from this, the unbiasedness of the estimator in (48) follows directly.

Statistical Properties of the OLS Estimator

The expectation or mean vector of $\hat{\beta}$, and its dispersion matrix as well, may be found from the expression

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'(X\beta + \varepsilon)\\ &= \beta + (X'X)^{-1}X'\varepsilon.\end{aligned} \tag{53}$$

The expectation is

$$\begin{aligned}E(\hat{\beta}) &= \beta + (X'X)^{-1}X'E(\varepsilon)\\ &= \beta.\end{aligned} \tag{54}$$

Thus, $\hat{\beta}$ is an unbiased estimator. The deviation of $\hat{\beta}$ from its expected value is $\hat{\beta} - E(\hat{\beta}) = (X'X)^{-1}X'\varepsilon$. Therefore, the dispersion matrix, which contains the variances and covariances of the elements of $\hat{\beta}$, is

$$\begin{aligned}D(\hat{\beta}) &= E\Big[\big\{\hat{\beta} - E(\hat{\beta})\big\}\big\{\hat{\beta} - E(\hat{\beta})\big\}'\Big]\\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}\\ &= \sigma^2(X'X)^{-1}.\end{aligned} \tag{55}$$

The Gauss–Markov theorem asserts that $\hat{\beta}$ is the unbiased linear estimator of least dispersion. Thus,

(56) If $\hat{\beta}$ is the OLS estimator of β, and if β^* is any other linear unbiased estimator of β, then $V(q'\beta^*) \geq V(q'\hat{\beta})$, where q is a constant vector.

Proof. Since $\beta^* = Ay$ is an unbiased estimator, it follows that $E(\beta^*) = AE(y) = AX\beta = \beta$, which implies that $AX = I$. Now write $A = (X'X)^{-1}X' + G$. Then, $AX = I$ implies that $GX = 0$. It follows that

$$\begin{aligned} D(\beta^*) &= AD(y)A' \\ &= \sigma^2\big\{(X'X)^{-1}X' + G\big\}\big\{X(X'X)^{-1} + G'\big\} \\ &= \sigma^2(X'X)^{-1} + \sigma^2 GG' = D(\hat{\beta}) + \sigma^2 GG'. \end{aligned} \tag{57}$$

Therefore, for any constant vector q of order k, there is

$$\begin{aligned} V(q'\beta^*) &= q'D(\hat{\beta})q + \sigma^2 q'GG'q \\ &\geq q'D(\hat{\beta})q = V(q'\hat{\beta}); \end{aligned} \tag{58}$$

and thus the inequality $V(q'\beta^*) \geq V(q'\hat{\beta})$ is established.

Orthogonality and Omitted-Variables Bias

Consider the partitioned regression model of equation (12), which was written as

$$(59)\qquad y = [\,X_1, X_2\,]\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

Imagine that the columns of X_1 are orthogonal to the columns of X_2 such that $X_1'X_2 = 0$.

In the partitioned form of the formula $\hat{\beta} = (X'X)^{-1}X'y$, there would be

$$(60)\quad X'X = \begin{bmatrix}X_1'\\ X_2'\end{bmatrix}[\,X_1 \quad X_2\,] = \begin{bmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{bmatrix} = \begin{bmatrix}X_1'X_1 & 0\\ 0 & X_2'X_2\end{bmatrix},$$

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

$$(61)\qquad (X'X)^{-1} = \begin{bmatrix}X_1'X_1 & 0\\ 0 & X_2'X_2\end{bmatrix}^{-1} = \begin{bmatrix}(X_1'X_1)^{-1} & 0\\ 0 & (X_2'X_2)^{-1}\end{bmatrix}.$$

There is also

$$X'y = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}. \tag{62}$$

On combining these elements, we find that

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1}X_1'y \\ (X_2'X_2)^{-1}X_2'y \end{bmatrix}. \tag{63}$$

In this 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 recognised that this result does not hold true in general. The general formulae for $\hat{\beta}_1$ and $\hat{\beta}_2$ are those that have been given already under (15) and (20):

$$\begin{aligned} \hat{\beta}_1 &= (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2), \\ \hat{\beta}_2 &= \left\{X_2'(I - P_1)X_2\right\}^{-1}X_2'(I - P_1)y, \quad P_1 = X_1(X_1'X_1)^{-1}X_1'. \end{aligned} \tag{64}$$

The purpose of including X_2 in the regression equation, when our 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.

If X_2 is erroneously excluded, then the estimate of β_1 will be

$$\begin{aligned}\tilde{\beta}_1 &= (X_1'X_1)^{-1}X_1'y \\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon) \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon.\end{aligned} \tag{65}$$

On applying the expectations operator, we find that

$$E(\tilde{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2, \tag{66}$$

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 estimator will be unbiased only when either $X_1'X_2 = 0$ or $\beta_2 = 0$. In other circumstances, it will suffer from *omitted-variables bias.*

Restricted Least-Squares Regression

A set of j linear restrictions on the vector β can be written as $R\beta = r$, where r is a $j \times k$ matrix of linearly independent rows, such that $\text{Rank}(R) = j$, and r is a vector of j elements.

To combine this *a priori* information with the sample information, the sum of squares $(y - X\beta)'(y - X\beta)$ is minimised subject to $R\beta = r$. This leads to the Lagrangean function

$$
\begin{aligned}
L &= (y - X\beta)'(y - X\beta) + 2\lambda'(R\beta - r) \\
&= y'y - 2y'X\beta + \beta'X'X\beta + 2\lambda'R\beta - 2\lambda'r.
\end{aligned} \tag{67}
$$

Differentiating L with respect to β and setting the result to zero, gives following first-order condition $\partial L/\partial\beta = 0$:

$$2\beta'X'X - 2y'X + 2\lambda'R = 0. \tag{68}$$

After transposing the expression, eliminating the factor 2 and rearranging, we have

$$X'X\beta + R'\lambda = X'y. \tag{69}$$

Combining these equations with the restrictions gives

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ r \end{bmatrix}. \tag{70}$$

For the system to given a unique value of β, the matrix $X'X$ need not be invertible—it is enough that the condition

$$\text{Rank} \begin{bmatrix} X \\ R \end{bmatrix} = k \tag{71}$$

should hold, which means that the matrix should have full column rank. Consider applying OLS to the equation

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} \varepsilon \\ 0 \end{bmatrix}, \tag{72}$$

which puts the equations of the observations and the equations of the restrictions on an equal footing. An estimator exits on the condition that $(X'X + R'R)^{-1}$ exists, for which the satisfaction of the rank condition is necessary and sufficient. Then, $\hat{\beta} = (X'X + R'R)^{-1}(X'y + R'r)$.

Let us assume that $(X'X)^{-1}$ *does* exist. Then equation (68) gives an expression for β in the form of

$$\begin{aligned}\beta^* &= (X'X)^{-1}X'y - (X'X)^{-1}R'\lambda \\ &= \hat{\beta} - (X'X)^{-1}R'\lambda,\end{aligned} \tag{73}$$

where $\hat{\beta}$ is the unrestricted ordinary least-squares estimator. Since $R\beta^* = r$, premultiplying the equation by R gives

$$r = R\hat{\beta} - R(X'X)^{-1}R'\lambda, \tag{74}$$

from which

$$\lambda = \{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r). \tag{75}$$

On substituting this expression back into equation (73), we get

$$\beta^* = \hat{\beta} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r). \tag{76}$$

This formula is an instance of the prediction-error algorithm, whereby the estimate of β is updated using information provided by the restrictions.

Given that $E(\hat{\beta} - \beta) = 0$, which is to say that $\hat{\beta}$ is an unbiased estimator, then, on the supposition that the restrictions are valid, it follows that $E(\beta^* - \beta) = 0$, so that β^* is also unbiased.

Next, consider the expression

$$
\begin{aligned}
\beta^* - \beta &= [I - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R](\hat{\beta} - \beta) \\
&= (I - P_R)(\hat{\beta} - \beta),
\end{aligned} \tag{77}
$$

where

$$
P_R = (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R. \tag{78}
$$

The expression comes from taking β from both sides of (76) and from recognising that $R\hat{\beta} - r = R(\hat{\beta} - \beta)$. It can be seen that P_R is an idempotent matrix that is subject to the conditions that

$$
P_R = P_R^2, \quad P_R(I - P_R) = 0 \quad \text{and} \quad P_R'X'X(I - P_R) = 0. \tag{79}
$$

From equation (77), it can be deduced that

$$
\begin{aligned}
D(\beta^*) &= (I - P_R)E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\}(I - P_R) \\
&= \sigma^2(I - P_R)(X'X)^{-1}(I - P_R) \\
&= \sigma^2[(X'X)^{-1} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}].
\end{aligned} \tag{80}
$$

Regressions on Trigonometrical Functions

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:

$$(81) \qquad y_t = \alpha_0 + \sum_{j=1}^{[T/2]} \{\alpha_j \cos(\omega_j t) + \beta_j \sin(\omega_j t)\}; \quad t = 0, 1, \ldots, T - 1.$$

The vectors of the generic trigonometric regressors may be denoted by

$$(83) \qquad c_j = [c_{0j}, c_{1j}, \ldots c_{T-1,j}]' \qquad \text{and} \qquad s_j = [s_{0j}, s_{1j}, \ldots s_{T-1,j}]',$$

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:

$$\begin{aligned} &c_i'c_j = s_i's_j = 0 \quad \text{if} \quad i \neq j, \\ &\text{and} \quad c_i's_j = 0 \quad \text{for all} \quad i, j. \end{aligned} \tag{84}$$

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

$$\begin{aligned} &c_0'c_0 = \iota'\iota = T, \qquad s_0's_0 = 0, \\ &c_j'c_j = s_j's_j = T/2 \quad \text{for} \quad j = 1, \ldots, [(T-1)/2] \end{aligned} \tag{85}$$

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

$$s_n's_n = 0, \qquad \text{and} \qquad c_n'c_n = T. \tag{86}$$

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

$$\alpha_0 = (\iota'\iota)^{-1}\iota' y = \frac{1}{T}\sum_t y_t = \bar{y}. \tag{87}$$

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

$$\alpha_j = (c_j' c_j)^{-1} c_j' y = \frac{2}{T}\sum_t y_t \cos\omega_j t, \tag{88}$$

and

$$\beta_j = (s_j' s_j)^{-1} s_j' y = \frac{2}{T}\sum_t y_t \sin\omega_j t. \tag{89}$$

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

$$\alpha_n = (c_n' c_n)^{-1} c_n' y = \frac{1}{T}\sum_t (-1)^t y_t. \tag{90}$$

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:

$$(91)\qquad y'y = \alpha_0^2 \iota'\iota + \sum_{j=1}^{[T/2]} \left\{ \alpha_j^2 c_j' c_j + \beta_j^2 s_j' s_j \right\}.$$

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

$$(92)\qquad (y - \bar{y})'(y - \bar{y}) = \frac{T}{2} \sum_{j=1}^{n-1} \left\{ \alpha_j^2 + \beta_j^2 \right\} + T\alpha_n^2 = \frac{T}{2} \sum_{j=1}^{n} \rho_j^2.$$

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

$$\frac{1}{T}\sum_{t=0}^{T-1}(y_t-\bar{y})^2=\frac{1}{2}\sum_{j=1}^{n}(\alpha_j^2+\beta_j^2). \tag{93}$$

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.

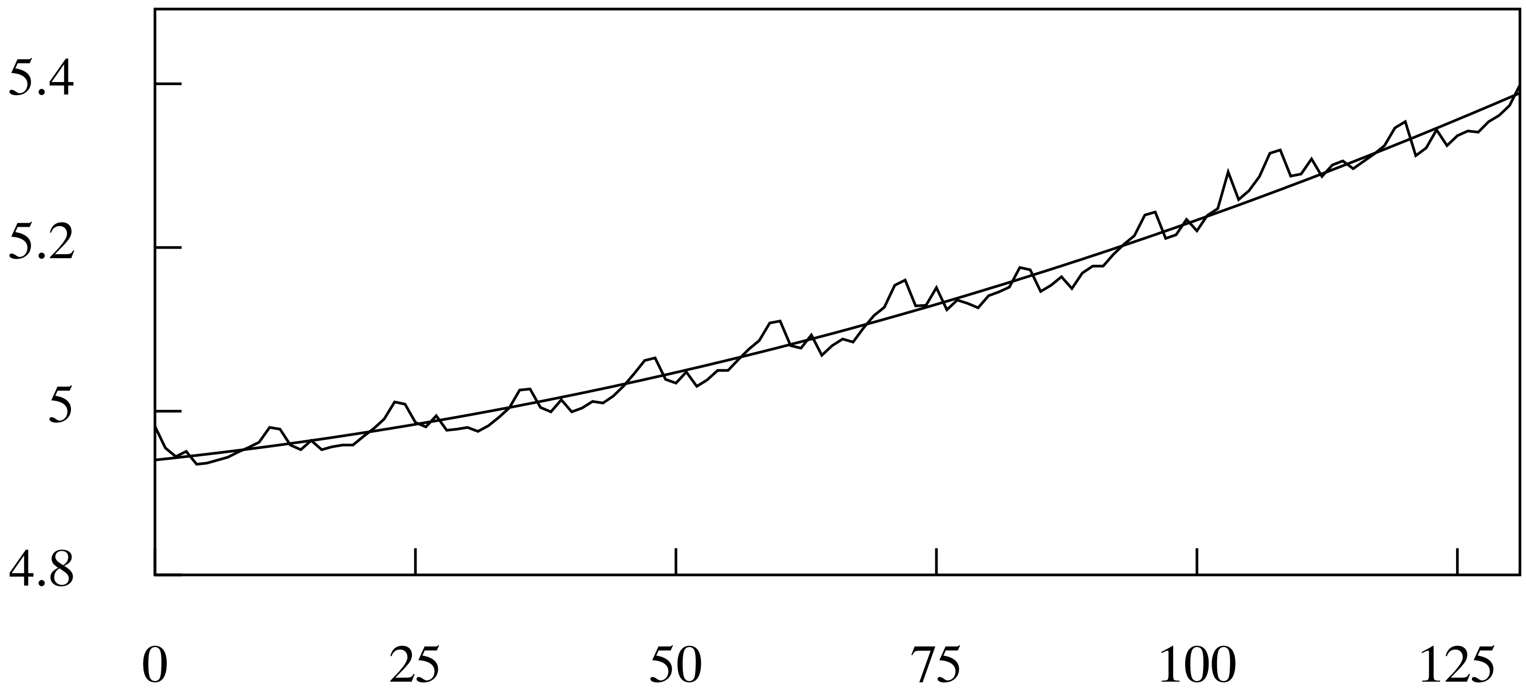

Figure 2. The plot of 132 monthly observations on the U.S. money supply, beginning in January 1960. A quadratic function has been interpolated through the data.

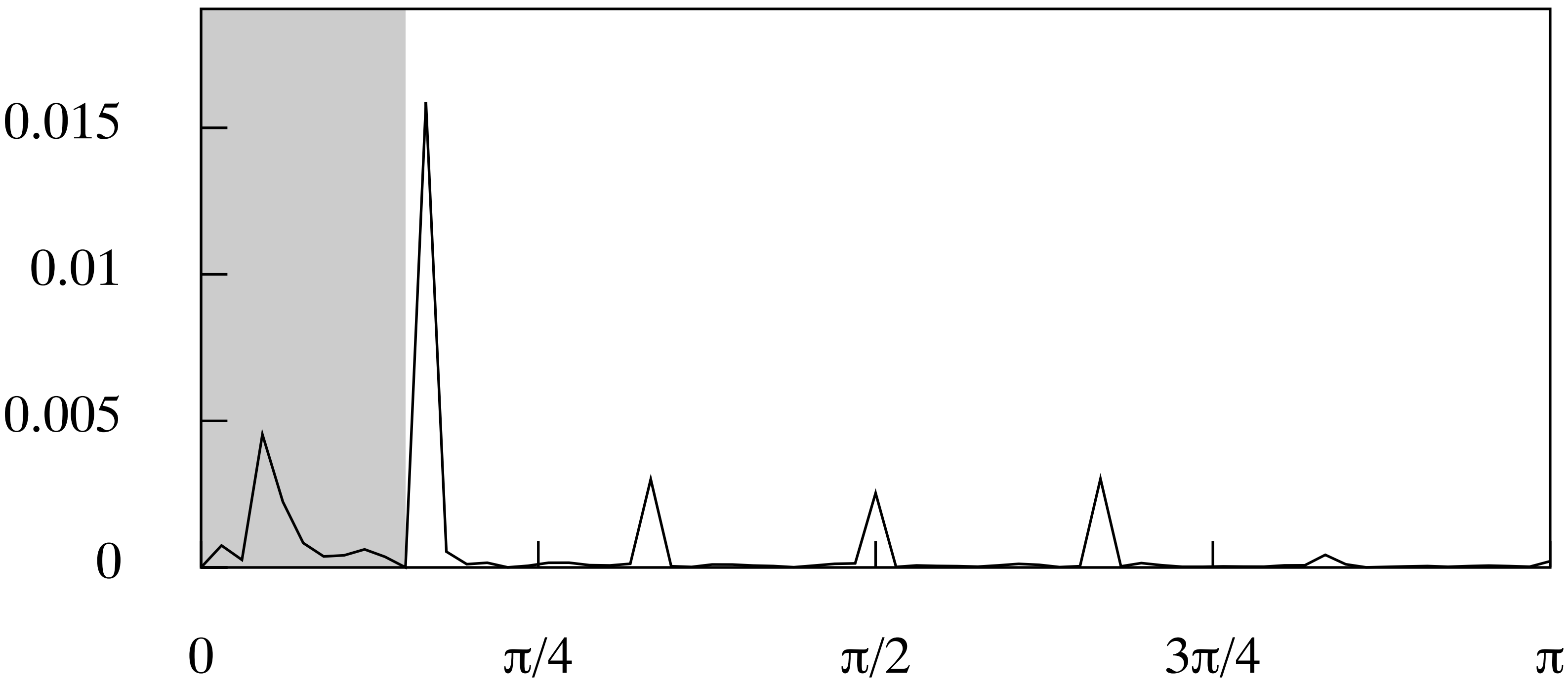

Figure 3. The periodogram of the residuals of the logarithmic money-supply data.