6 : CHAPTER

The Classical Linear Regression Model

In this lecture, we shall present the basic theory of the classical statistical method of regression analysis.

The Linear Regression Model

A regression equation of the form

$$
(1) \qquad \begin{aligned} y_t &= x_{t1}\beta_1 + x_{t2}\beta_2 + \cdots + x_{tk}\beta_k + \varepsilon_t \\ &= x_{t.}\beta + \varepsilon_t \end{aligned}
$$

explains the value of a dependent variable y_t in terms of a set of k observable variables in $x_{t.} = [x_{t1}, x_{t2}, \ldots, x_{tk}]$ and an unobservable random variable ε_t. The vector $\beta = [\beta_1, \beta_2, \ldots, \beta_k]'$ contains the parameters of a linear combination of the variables in x_t. A set of T successive realisations of the regression relationship, indexed by $t = 1, 2, \ldots, T$, can be compiled into a system

$$
(2) \qquad y = X\beta + \varepsilon,
$$

wherein $y = [y_1, y_2, \ldots, y_T]'$ and $\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T]'$ are vectors of order T and $X = [x_{tk}]$ is a matrix of order $T \times k$. We shall assume that X is a non-stochastic matrix with $\text{Rank}(X) = k$ which requires that $T \geq k$.

According to the classical assumptions, the elements of the disturbance vector ε are distributed independently and identically with expected values of zero and a common variance of σ^2. Thus,

$$
(3) \qquad E(\varepsilon) = 0 \qquad \text{and} \qquad D(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I_T.
$$

The matrix $D(\varepsilon)$, which is described as the variance–covariance matrix or the dispersion matrix of ε, contains the common variance $\sigma^2 = E[\{\varepsilon_t - E(\varepsilon_t)\}^2]$ in each of its diagonal locations. Its other locations contain zero-valued elements, each of which corresponds to the covariance $E[\{\varepsilon_t - E(\varepsilon_t)\}\{\varepsilon_s - E(\varepsilon_s)\}']$ of two distinct elements of ε.

The value of β may estimated according to the principle of ordinary least-squares regression by minimising the quadratic function

$$S = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta). \tag{4}$$

The problem can be envisaged as one of finding a value for $\mu = X\beta$ residing, at a minimum distance from the vector y, in the subspace or the manifold spanned by the columns of X. This interpretation comes from recognising that the function $S = (y - X\beta)'(y - X\beta)$ represents the square of the Euclidean distance between the two vectors.

The minimising value of β is found by differentiating the function $S(\beta)$ with respect to β and setting the result to zero. This gives the condition

$$\frac{\partial S}{\partial \beta} = 2\beta'X'X - 2y'X = 0. \tag{5}$$

By rearranging the condition, the so-called normal equations are obtained

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

whose solution is the ordinary least-squares estimate of the regression parameters:

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

The estimate of the systematic component of the regression equations is

$$\begin{aligned} X\hat{\beta} &= X(X'X)^{-1}X'y \\ &= Py. \end{aligned} \tag{8}$$

Here $P = X(X'X)^{-1}X'$, which is called the orthogonal or perpendicular projector on the manifold of X, is a symmetric idempotent matrix with the properties that $P = P' = P^2$.

The Decomposition of the Sum of Squares

Ordinary least-squares regression entails the decomposition the vector y into two mutually orthogonal components. These are the vector $Py = X\hat{\beta}$, which estimates the systematic component of the regression equation, and the residual vector $e =$y$-X\hat{\beta}$, which estimates the disturbance vector ε. The condition that e should be orthogonal to the manifold of X in which the systematic component resides, such that $X'e = X'(y - X\hat{\beta}) = 0$, is precisely the condition which is expressed by the normal equations (6).

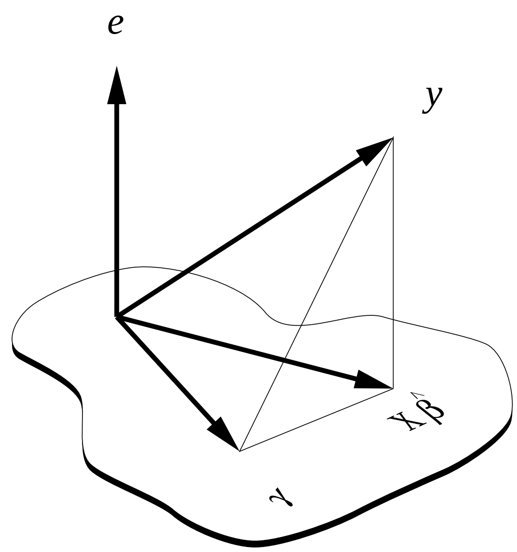

Figure 1. The vector $Py = X\hat{\beta}$ is formed by the orthogonal projection of the vector y onto the subspace spanned by the columns of the matrix X.

Corresponding to the decomposition of y, there is a decomposition of the sum of squares $y'y$. To express the latter, let us write $X\hat{\beta} = Py$ and $e = y - X\hat{\beta} = (I - P)y$. Then, in consequence of the condition $P = P' = P^2$ and the equivalent condition $P'(I - P) = 0$, it follows that

$$
\begin{aligned}
y'y &= \big\{Py + (I-P)y\big\}'\big\{Py + (I-P)y\big\} \\
&= y'Py + y'(I-P)y \\
&= \hat{\beta}'X'X\hat{\beta} + e'e.
\end{aligned} \tag{9}
$$

This is simply an instance of Pythagoras theorem; and the identity is expressed by saying 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 $e'e$. A geometric interpretation of the orthogonal decomposition of y and of the resulting Pythagorean relationship is given in Figure 1.

It is clear from intuition that, by projecting y perpendicularly onto the manifold of X, the distance between y and $Py = X\hat{\beta}$ is minimised. In order to establish this point formally, imagine that $\gamma = Pg$ is an arbitrary vector in the manifold of X. Then the Euclidean distance from y to γ cannot be less than the distance from y to $X\hat{\beta}$. The square of the former distance is

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

The properties of the projector P which have been used in simplifying equation (9), indicate that

$$(11)\qquad \begin{aligned}(y-\gamma)'(y-\gamma) &= y'(I-P)y + (y-g)'P(y-g)\\ &= e'e + (X\hat{\beta}-\gamma)'(X\hat{\beta}-\gamma).\end{aligned}$$

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 $e = y - X\hat{\beta}$; and this proves the assertion.

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

$$(12)\qquad \begin{aligned}R^2 &= \frac{\hat{\beta}'X'X\hat{\beta}}{y'y}\\ &= \frac{y'Py}{y'y}.\end{aligned}$$

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 $+1$.

Some Statistical Properties of the Estimator

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

$$(13)\qquad \begin{aligned}\hat{\beta} &= (X'X)^{-1}X'(X\beta+\varepsilon)\\ &= \beta + (X'X)^{-1}X'\varepsilon.\end{aligned}$$

On the assumption that the elements of X are nonstochastic, the expectation is given by

$$(14)\qquad \begin{aligned}E(\hat{\beta}) &= \beta + (X'X)^{-1}X'E(\varepsilon)\\ &= \beta.\end{aligned}$$

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

$$(15)\qquad \begin{aligned}D(\hat{\beta}) &= E\Big[\big\{\hat{\beta}-E(\hat{\beta})\big\}\big\{\hat{\beta}-E(\hat{\beta})\big\}'\Big]\\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}\\ &= \sigma^2(X'X)^{-1}.\end{aligned}$$

The Gauss–Markov theorem asserts that $\hat{\beta}$ is the unbiased linear estimator of least dispersion. This dispersion is usually characterised in terms of the variance of an arbitrary linear combination of the elements of $\hat{\beta}$, although it may also be characterised in terms of the determinant of the dispersion matrix $D(\hat{\beta})$. Thus,

(16) If $\hat{\beta}$ is the ordinary least-squares estimator of β in the classical linear regression model, and if β^* is any other linear unbiased estimator of β, then $V(q'\beta^*) \geq V(q'\hat{\beta})$ where q is any constant vector of the appropriate order.

Proof. Since $\beta^* = Ay$ is an unbiased estimator, it follows that $E(\beta^*) = AE(y) = AX\beta = \beta$, which implies that $AX = I$. Now set $A = (X'X)^{-1}X' + G$. Then $AX = I$ implies that $GX = 0$. Given that $D(y) = D(\varepsilon) = \sigma^2 I$, it follows that

$$
\begin{aligned}
D(\beta^*) &= AD(y)A' \\
&= \sigma^2\big\{(X'X)^{-1}X' + G\big\}\big\{X(X'X)^{-1} + G'\big\} \\
&= \sigma^2(X'X)^{-1} + \sigma^2 GG' \\
&= D(\hat{\beta}) + \sigma^2 GG'.
\end{aligned} \tag{17}
$$

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

$$
\begin{aligned}
V(q'\beta^*) &= q'D(\hat{\beta})q + \sigma^2 q'GG'q \\
&\geq q'D(\hat{\beta})q = V(q'\hat{\beta});
\end{aligned} \tag{18}
$$

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

Estimating the Variance of the Disturbance

The principle of least squares does not, of its own, suggest a means of estimating the disturbance variance $\sigma^2 = V(\varepsilon_t)$. However, it is natural to estimate the moments of a probability distribution by their empirical counterparts. Given that $e_t = y - x_{t.}\hat{\beta}$ is an estimate 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

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{T-k}\sum_{t=1}^{T} e_t^2 \\
&= \frac{1}{T-k}(y - X\hat{\beta})'(y - X\hat{\beta}).
\end{aligned} \tag{19}
$$

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

$$E\big\{(y - X\hat{\beta})'(y - X\hat{\beta})\big\} = E(\varepsilon'\varepsilon) - E(\varepsilon' P\varepsilon). \tag{20}$$

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

$$E(\varepsilon'\varepsilon) = \sum_{t=1}^{T} E(e_t^2) = T\sigma^2. \tag{21}$$

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

$$\begin{aligned} E(\varepsilon' P\varepsilon) &= \mathrm{Trace}\big\{E(\varepsilon' P\varepsilon)\big\} = E\big\{\mathrm{Trace}(\varepsilon' P\varepsilon)\big\} = E\big\{\mathrm{Trace}(\varepsilon\varepsilon' P)\big\} \\ &= \mathrm{Trace}\big\{E(\varepsilon\varepsilon')P\big\} = \mathrm{Trace}\big\{\sigma^2 P\big\} = \sigma^2\mathrm{Trace}(P) \\ &= \sigma^2 k. \end{aligned} \tag{22}$$

The final equality follows from the fact that $\mathrm{Trace}(P) = \mathrm{Trace}(I_k) = k$. Putting the results of (21) and (22) into (20), gives

$$E\big\{(y - X\hat{\beta})'(y - X\hat{\beta})\big\} = \sigma^2(T - k); \tag{23}$$

and, from this, the unbiasedness of the estimator in (19) follows directly.

The Partitioned Regression Model

In testing hypotheses, it is helpful to have explicit expressions for the subvectors within $\hat{\beta}' = [\hat{\beta}_1', \hat{\beta}_2']$. To this end, the equations of (2) may be written as $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, where X_1 and X_2 contain T observations on k_1 and k_2 variables respectively. The normal equations of (6) may be partitioned conformably to give

$$\begin{aligned} X_1'X_1\beta_1 + X_1'X_2\beta_2 &= X_1'y \qquad \text{and} \\ X_2'X_1\beta_1 + X_2'X_2\beta_2 &= X_2'y. \end{aligned} \tag{24}$$

Premultiplying the first of these by $X_2'X_1(X_1'X_1)^{-1}$ and subtracting it from the second gives

$$\big\{X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2\big\}\beta_2 = X_2'y - X_2'X_1(X_1'X_1)^{-1}X_1'y. \tag{25}$$

When the projector $P_1 = X_1(X_1'X_1)^{-1}X_1'$ is defined, the equation may be written more intelligibly, as $X_2'(I - P_1)X_2\beta_2 = X_2'(I - P_1)y$. The estimate of β_2 is given by

$$\hat{\beta}_2 = \big\{X_2'(I - P_1)X_2\big\}^{-1}X_2'(I - P_1)y. \tag{26}$$

An analogous expression is available for $\hat{\beta}_1$. However, knowing the value of $\hat{\beta}_2$ enables us to obtain $\hat{\beta}_1$ alternatively from the expression

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2) \tag{27}$$

which comes directly from the first equation of (24).

Some Matrix Identities

The estimators of β_1 and β_2 may also be derived by using the partitioned form of the matrix $(X'X)^{-1}$. This is given by

$$\begin{aligned} &\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \bigl\{X_1'(I - P_2)X_1\bigr\}^{-1} & -\bigl\{X_1'(I - P_2)X_1\bigr\}^{-1}X_1'X_2(X_2'X_2)^{-1} \\ -\bigl\{X_2'(I - P_1)X_2\bigr\}^{-1}X_2'X_1(X_1'X_1)^{-1} & \bigl\{X_2'(I - P_1)X_2\bigr\}^{-1} \end{bmatrix} \end{aligned} \tag{28}$$

The result is easily verified by postmultiplying the matrix on the RHS by the partitioned form of $X'X$ to give a partitioned form of the identity matrix.

By forming the projector $P = X(X'X)^{-1}X'$ from $X = [X_1, X_2]$ and from the partitioned form of $(X'X)^{-1}$, it may be shown that

$$\begin{aligned} P &= P_{1/2} + P_{2/1}, \qquad \text{where} \\ P_{1/2} &= X_1\bigl\{X_1'(I - P_2)X_1\bigr\}^{-1}X_1'(I - P_2) \qquad \text{and} \\ P_{2/1} &= X_2\bigl\{X_2'(I - P_1)X_2\bigr\}^{-1}X_2'(I - P_1). \end{aligned} \tag{29}$$

In the notation of the regression model, the identity $Py = P_{1/2}y + P_{2/1}y$ is expressed as $X\hat{\beta} = X_1\hat{\beta}_1 + X_2\hat{\beta}_2$.

The restriction of the transformation $P_{1/2}$ to the manifold of X may be described as the oblique projection onto the manifold of X_1 along the manifold of X_2. This means that the manifold of X_2 falls within the null space of the projector. The corresponding conditions $P_{1/2}X_1 = X_1$ and $P_{1/2}X_2 = 0$ are readily confirmed. Thus,

$$\begin{aligned} P_{1/2}P_1 &= P_1, \\ P_{1/2}P_2 &= 0. \end{aligned} \tag{30}$$

Likewise, $P_{2/1}X_2 = X_2$ and $P_{2/1}X_1 = 0$. These conditions indicate that

$$\begin{aligned} PP_1 &= (P_{1/2} + P_{2/1})P_1 \\ &= P_1 \\ &= P_1P. \end{aligned} \tag{31}$$

The final equality follows from the symmetry of P_1 and P.

Now consider premultiplying and postmultiplying the partitioned form of $(X'X)^{-1}$ by $(I - P_1)X = [0, (I - P_1)X_2]$ and its transpose respectively. This gives

$$\begin{aligned}(I - P_1)X(X'X)^{-1}X'(I - P_1) &= (I - P_1)P(I - P_1)\\ &= (I - P_1)X_2\big\{X_2'(I - P_1)X_2\big\}^{-1}X_2'(I - P_1).\end{aligned} \tag{32}$$

But the conditions $PP_1 = P_1P = P_1$ can be used to show that $(I - P_1)P(I - P_1) = P - P_1$. Thus, an important identity is derived in the form of

$$(I - P_1)X_2\big\{X_2'(I - P_1)X_2\big\}^{-1}X_2'(I - P_1) = P - P_1. \tag{33}$$

This will be used in the sequel.

The Normal Distribution and the Sampling Distributions

It is often appropriate to assume that the elements of the disturbance vector ε within the regression equations $y = X\beta + \varepsilon$ are distributed independently and identically according to a normal law. Under this assumption, the sampling distributions of the estimates may be derived and various hypotheses relating to the underlying parameters may be tested.

To denote that x is a normally distributed random variable with a mean of $E(x) = \mu$ and a dispersion matrix of $D(x) = \Sigma$, we shall write $x \sim N(\mu, \Sigma)$. A vector $z \sim N(0, I)$ with a mean of zero and a dispersion matrix of $D(z) = I$ is described as a standard normal vector. Any normal vector $x \sim N(\mu, \Sigma)$ can be standardised:

(34) If T is a transformation such that $T\Sigma T' = I$ and $T'T = \Sigma^{-1}$, then $T(x - \mu) \sim N(0, I)$.

Associated with the normal distribution are a variety of so-called sampling distributions which occur frequently in problems of statistical inference. Amongst these are the chi-square distribution, the F distribution and the t distribution.

If $z \sim N(0, I)$ is a standard normal vector of n elements, then the sum of squares of its elements has a chi-square distribution of n degrees of freedom; and this is denoted by $z'z \sim \chi^2(n)$. With the help of the standardising transformation, it can be shown that,

(35) If $x \sim N(\mu, \Sigma)$ is a vector of order n, then $(x - \mu)'\Sigma^{-1}(x - \mu) \sim \chi^2(n)$.

The sum of any two independent chi-square variates is itself a chi-square variate whose degrees of freedom equal the sum of the degrees of freedom of its constituents. Thus,

(36) If $u \sim \chi^2(m)$ and $v \sim \chi^2(n)$ are independent chi-square variates of m and n degrees of freedom respectively, then $(u+v) \sim \chi^2(m+n)$ is a chi-square variate of $m+n$ degrees of freedom.

The ratio of two independent chi-square variates divided by their respective degrees of freedom has a F distribution which is completely characterised by these degrees of freedom. Thus,

(37) If $u \sim \chi^2(m)$ and $v \sim \chi^2(n)$ are independent chi-square variates, then the variate $F = (u/m)/(v/n)$ has an F distribution of m and n degrees of freedom; and this is denoted by writing $F \sim F(m, n)$.

The sampling distribution which is most frequently used is the t distribution. A t variate is a ratio of a standard normal variate and the root of an independent chi-square variate divided by its degrees of freedom. Thus,

(38) If $z \sim N(0, 1)$ and $v \sim \chi^2(n)$ are independent variates, then $t = z/\sqrt{(v/n)}$ has a t distribution of n degrees of freedom; and this is denoted by writing $t \sim t(n)$.

It is clear that $t^2 \sim F(1, n)$.

Hypothesis Concerning the Coefficients

A linear function of a normally distributed vector is itself normally distributed. Thus, it follows that, if $y \sim N(X\beta, \sigma^2 I)$, then

$$\hat{\beta} \sim N_k\{\beta, \sigma^2(X'X)^{-1}\}. \tag{39}$$

Likewise, the marginal distributions of $\hat{\beta}_1, \hat{\beta}_2$ within $\hat{\beta}' = [\hat{\beta}_1, \hat{\beta}_2]$ are given by

$$\hat{\beta}_1 \sim N_{k_1}\big(\beta_1, \sigma^2\{X_1'(I - P_2)X_1\}^{-1}\big), \tag{40}$$

$$\hat{\beta}_2 \sim N_{k_2}\big(\beta_2, \sigma^2\{X_2'(I - P_1)X_2\}^{-1}\big). \tag{41}$$

On applying the result under (35) to (39), we find that

$$\sigma^{-2}(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \sim \chi^2(k). \tag{42}$$

Similarly, it follows from (40) and (41) that

$$\sigma^{-2}(\hat{\beta}_1 - \beta_1)'X_1'(I - P_2)X_1(\hat{\beta}_1 - \beta_1) \sim \chi^2(k_1), \tag{43}$$

$$\sigma^{-2}(\hat{\beta}_2 - \beta_2)'X_2'(I - P_1)X_2(\hat{\beta}_2 - \beta_2) \sim \chi^2(k_2). \tag{44}$$

The distribution of the residual vector $e = y - X\hat{\beta}$ is degenerate in the sense that the mapping $e = (I - P)\varepsilon$ from the disturbance vector ε to the

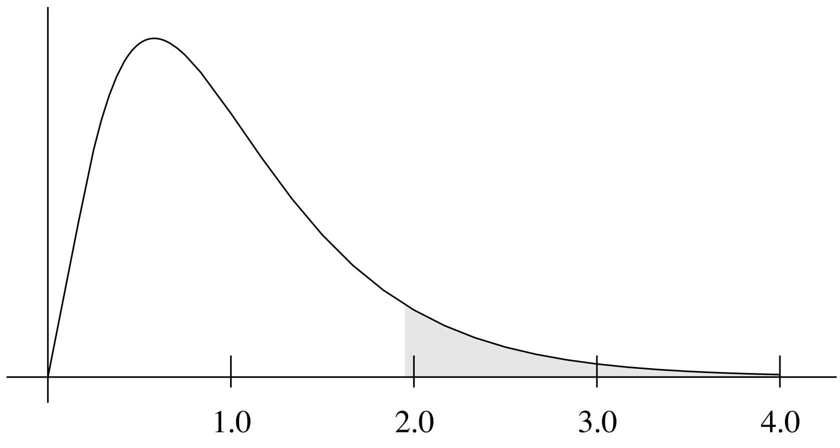

Figure 2. The critical region, at the 10% significance level, of an $F(5, 60)$ statistic.

residual vector e entails a singular transformation. Nevertheless, it is possible to obtain a factorisation of the transformation in the form of $I - P = CC'$, where C is matrix of order $T \times (T-k)$ comprising $T-k$ orthonormal columns which are orthogonal to the columns of X such that $C'X = 0$. Now $C'C = I_{T-k}$; so it follows that, on premultiplying $y \sim N_T(X\beta, \sigma^2 I)$ by C', we get $C'y \sim N_{T-k}(0, \sigma^2 I)$. Hence

$$\sigma^{-2} y'CC'y = \sigma^{-2}(y - X\hat{\beta})'(y - X\hat{\beta}) \sim \chi^2(T-k). \tag{45}$$

The vectors $X\hat{\beta} = Py$ and $y - X\hat{\beta} = (I-P)y$ have a zero-valued covariance matrix. If two normally distributed random vectors have a zero covariance matrix, then they are statistically independent. Therefore it follows that

$$\begin{aligned} &\sigma^{-2}(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \sim \chi^2(k) \quad \text{and} \\ &\sigma^{-2}(y - X\hat{\beta})'(y - X\hat{\beta}) \sim \chi^2(T-k) \end{aligned} \tag{46}$$

are mutually independent chi-square variates. From this, it can be deduced that

$$\begin{aligned} F &= \left\{ \frac{(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)}{k} \Bigg/ \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T-k} \right\} \\ &= \frac{1}{\hat{\sigma}^2 k}(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \sim F(k, T-k). \end{aligned} \tag{47}$$

To test an hypothesis specifying that $\beta = \beta_\diamond$, we simply insert this value in the above statistic and compare the resulting value with the critical values of an F

distribution of k and $T-k$ degrees of freedom. If a critical value is exceeded, then the hypothesis is liable to be rejected.

The test is readily intelligible since it is based on a measure of the distance between the hypothesised value $X\beta_\diamond$ of the systematic component of the regression and the value $X\hat{\beta}$ which is suggested by the data. If the two values are remote from each other, then we may suspect that the hypothesis is at fault.

It is usual to suppose that a subset of the elements of the parameter vector β are zeros. This represents an instance of a class of hypotheses which specify values for a subvector β_2 within the partitioned model $y = X_1\beta_1 + X\beta_2 + \varepsilon$ without asserting anything about the values of the remaining elements in the subvector β_1. The appropriate test statistic for testing the hypothesis that $\beta_2 = \beta_{2\diamond}$ is

$$F = \frac{1}{\hat{\sigma}^2 k_2}(\hat{\beta}_2 - \beta_{2\diamond})' X_2'(I - P_1)X_2(\hat{\beta}_2 - \beta_{2\diamond}). \tag{48}$$

This will have an $F(k_2, T-k)$ distribution if the hypothesis is true.

A limiting case of the F statistic concerns the test of an hypothesis affecting a single element β_i within the vector β. By specialising the expression under (48), a statistic may be derived in the form of

$$F = \frac{(\hat{\beta}_i - \beta_{i\diamond})^2}{\hat{\sigma}^2 w_{ii}}, \tag{49}$$

wherein w_{ii} stands for the ith diagonal element of $(X'X)^{-1}$. If the hypothesis is true, then this will be distributed according to the $F(1, T-k)$ law. However, the usual way of assessing such an hypothesis is to relate the value of the statistic

$$t = \frac{\hat{\beta}_i - \beta_{i\diamond}}{\sqrt{(\hat{\sigma}^2 w_{ii})}} \tag{50}$$

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

Cochrane's Theorem and the Decomposition of a Chi-Square Variate

The standard test of an hypothesis regarding the vector β in the model $N(y; X\beta, \sigma^2 I)$ entails a multi-dimensional version of Pythagoras' Theorem. Consider the decomposition of the vector y into the systematic component and the residual vector. This gives

$$\begin{gathered} y = X\hat{\beta} + (y - X\hat{\beta}) \quad \text{and} \\ y - X\beta = (X\hat{\beta} - X\beta) + (y - X\hat{\beta}), \end{gathered} \tag{51}$$

where the second equation comes from subtracting the unknown mean vector $X\beta$ from both sides of the first. These equations can also be expressed in terms of the projector $P = X(X'X)^{-1}X'$ which gives $Py = X\hat{\beta}$ and $(I - P)y = y - X\hat{\beta} = e$. Also, the definition $\varepsilon = y - X\beta$ can be used within the second of the equations. Thus,

$$(52) \qquad \begin{aligned} y &= Py + (I - P)y \qquad \text{and} \\ \varepsilon &= P\varepsilon + (I - P)\varepsilon. \end{aligned}$$

The reason for adopting this notation is that it enables us to envisage more clearly the Pythagorean relationship between the vectors. Thus, from the fact that $P = P' = P^2$ and that $P'(I - P) = 0$, it can be established that

$$(53) \qquad \begin{aligned} \varepsilon'\varepsilon &= \varepsilon'P\varepsilon + \varepsilon'(I - P)\varepsilon \qquad \text{or, equivalently,} \\ \varepsilon'\varepsilon &= (X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta) + (y - X\hat{\beta})'(y - X\hat{\beta}). \end{aligned}$$

The terms in these expressions represent squared lengths; and the vectors themselves form the sides of a right-angled triangle with $P\varepsilon$ at the base, $(I - P)\varepsilon$ as the vertical side and ε as the hypotenuse. These relationship are represented by Figure 1 where $\gamma = X\beta$ and where $\varepsilon = y - \gamma$.

The usual test of an hypothesis regarding the elements of the vector β is based on the foregoing relationships. Imagine that the hypothesis postulates that the true value of the parameter vector is $\beta_\diamond$. To test this proposition, we compare the value of $X\beta_\diamond$ with the estimated mean vector $X\hat{\beta}$. The test is a matter of assessing the proximity of the two vectors which is measured by the square of the distance which separates them. This would be given by

$$(54) \qquad \varepsilon'P\varepsilon = (X\hat{\beta} - X\beta_\diamond)'(X\hat{\beta} - X\beta_\diamond)$$

If the hypothesis is untrue and if $X\beta_\diamond$ is remote from the true value of $X\beta$, then the distance is liable to be excessive.

The distance can only be assessed in comparison with the variance σ^2 of the disturbance term or with an estimate thereof. Usually, one has to make do with the estimate of σ^2 which is provided by

$$(55) \qquad \begin{aligned} \hat{\sigma}^2 &= \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T - k} \\ &= \frac{\varepsilon'(I - P)\varepsilon}{T - k}. \end{aligned}$$

The numerator of this estimate is simply the squared length of the vector $e = (I - P)y = (I - P)\varepsilon$ which constitutes the vertical side of the right-angled triangle.

Simple arguments, which have been given in the previous section, serve to demonstrate that

$$
(56)\qquad
\begin{aligned}
&\text{(a)}\quad \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) \sim \sigma^2\chi^2(T),\\
&\text{(b)}\quad \varepsilon'P\varepsilon = (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \sim \sigma^2\chi^2(k),\\
&\text{(c)}\quad \varepsilon'(I - P)\varepsilon = (y - X\hat{\beta})'(y - X\hat{\beta}) \sim \sigma^2\chi^2(T - k),
\end{aligned}
$$

where (b) and (c) represent statistically independent random variables whose sum is the random variable of (a). These quadratic forms, divided by their respective degrees of freedom, find their way into the F statistic of (47) which is

$$
(57)\qquad F = \left\{ \frac{\varepsilon'P\varepsilon}{k} \bigg/ \frac{\varepsilon'(I - P)\varepsilon}{T - k} \right\} \sim F(k, T - k).
$$

A more elaborate decomposition of the $\chi^2(T)$ variate that the one above is often called for. In such cases, we can invoke Cochrane's theorem of which the following is a general statement.

(58) Let $\eta \sim N(0, I_n)$, and let $P = \sum P_i$ be a sum of k symmetric matrices with $\text{rank}(P) = r$ and $\text{rank}(P_i) = r_i$ such that $P_i = P_i^2$ and $P_iP_j = 0$ when $i \neq j$. Then $\eta'P_i\eta \sim \chi^2(r_i); i = 1, \ldots, k$ are independent chi-square variates such that $\sum \eta'P_i\eta = \eta'P\eta \sim \chi^2(r)$ with $r = \sum r_i$.

Proof. If the conditions of the theorem are satisfied, then there exists a partitioned $n \times r$ matrix of orthonormal vectors $C = [C_1, \ldots, C_k]$ such that $C'C = I$, $C_i'C_j = 0$ and $C_iC_i' = P_i$. If $\eta \sim N_n(0, I)$, then $C'\eta \sim N_r(0, I)$; and this can be written as

$$
C'\eta = \begin{bmatrix} C_1'\eta \\ C_2'\eta \\ \vdots \\ C_k'\eta \end{bmatrix} \sim N_r\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} I_{r_1} & 0 & \ldots & 0 \\ 0 & I_{r_2} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & I_{r_k} \end{bmatrix} \right),
$$

wherein $C_i'\eta \sim N_{r_i}(0, I)$ for $i = 1, \ldots, k$ are mutually independent standard normal variates. Thus, $\eta'CC'\eta \sim \chi^2(r)$ is a chi-square variate and also $\eta'C_iC_i'\eta \sim \chi^2(r_i)$ for $i = 1, \ldots, k$ constitute a set of mutually independent chi-square variates. Now observe that $\eta'CC'\eta = \eta'[C_1C_1' + \cdots + C_kC_k']\eta = \sum \eta'C_iC_i'\eta$. Thus, using $P_i = C_iC_i'$ and the notation $P = CC'$, we have $\sum \eta'P_i\eta = \eta'P\eta \sim \chi^2(r)$. Finally, it is clear from the construction that $r = \sum r_i$.

For an immediate application of the theorem, let $P = X(X'X)^{-1}X'$ where $X = [X_1, X_2]$ and let $P_1 = X_1(X_1'X_1)^{-1}X_1'$. Then consider the following decomposition:

$$\varepsilon = (I - P)\varepsilon + (P - P_1)\varepsilon + P_1\varepsilon. \tag{59}$$

Here the symmetric idempotent matrices $I - P$, $P - P_1$ and P_1 are mutually orthogonal. It follow that

$$\varepsilon'\varepsilon = \varepsilon'(I - P)\varepsilon + \varepsilon'(P - P_1)\varepsilon + \varepsilon' P_1\varepsilon. \tag{60}$$

Moreover, if $y - X\beta = \varepsilon \sim N(0, \sigma^2 I)$, then, according to Cochrane's theorem, we should have

$$\begin{aligned}
&\text{(a)} \quad && \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) \sim \sigma^2\chi^2(T),\\
&\text{(b)} \quad && \varepsilon'(I - P)\varepsilon = (y - X\hat{\beta})'(y - X\hat{\beta}) \sim \sigma^2\chi^2(T - k),\\
&\text{(c)} \quad && \varepsilon'(P - P_1)\varepsilon = (\hat{\beta}_2 - \beta_2)'X_2'(I - P_1)X_2(\hat{\beta}_2 - \beta_2) \sim \sigma^2\chi^2(k_2),\\
&\text{(d)} \quad && \varepsilon' P_1\varepsilon = (y - X\beta)'P_1(y - X\beta) \sim \sigma^2\chi^2(k_1),
\end{aligned} \tag{61}$$

where (b), (c) and (d) represent statistically independent random variables whose sum is the random variable of (a).

To obtain the result under (c), we may observe that $P - P_1 = (I - P_1)P$ and that $P\varepsilon = X\hat{\beta} - X\beta$. Then it can be seen that

$$\begin{aligned}
(P - P_1)\varepsilon &= (I - P_1)(X\hat{\beta} - X\beta)\\
&= (I - P_1)(X_2\hat{\beta}_2 - X_2\beta_2),
\end{aligned} \tag{62}$$

where the final equality follows from the fact that $(I - P_1)X_1 = 0$. The result follows in view of the symmetry and idempotency of $I - P_1$.

These quadratic forms under (b) and (c), divided by their respective degrees of freedom, find their way into the F statistic of (48) which is

$$F = \left\{ \frac{\varepsilon'(P - P_1)\varepsilon}{k_2} \Big/ \frac{\varepsilon'(I - P)\varepsilon}{T - k} \right\} \sim F(k_2, T - k). \tag{63}$$

An Alternative Formulation of the F statistic

An alternative way of forming the F statistic uses the products of two separate regressions. Consider the identity

$$\varepsilon'(P - P_1)\varepsilon = \varepsilon'(I - P_1)\varepsilon - \varepsilon'(I - P)\varepsilon. \tag{64}$$

The term of the LHS is the quadratic product which appears in the numerator of the F statistic of (48) and (63). The first term on the RHS can be written as

$$\begin{aligned} \varepsilon'(I-P_1)\varepsilon &= (y-X\beta)'(I-P_1)(y-X\beta) \\ &= (y-X_2\beta_2)'(I-P_1)(y-X_2\beta_2). \end{aligned} \tag{65}$$

Under the hypothesis that $\beta_2=\beta_{2\diamond}$, the term amounts to the residual sum of squares from the regression of $y-X_2\beta_{2\diamond}$ on X_1. This is the regression which comes from substituting the hypothesised value of $X_2\beta_2$ into the first of the normal equations of the partitioned regression model which are given under (24). The resulting regression equation is in the form of (27) with $\beta_{2\diamond}$ in place of $\hat{\beta}_2$.

The residual sum of squares of (65) may be described as the restricted sum of squares and denoted by RSS. The second term on the RHS of (64) is just the ordinary residual sum of squares

$$\begin{aligned} \varepsilon'(I-P)\varepsilon &= (y-X\beta)'(I-P)(y-X\beta) \\ &= y'(I-P)y. \end{aligned} \tag{66}$$

This may be obtained, equally, from the regression of y on X or from the regression of $y-X_2\beta_{2\diamond}$ on X; and it may be described as the unrestricted residual sum of squares and denoted by USS. From these considerations, it follows that the statistic for testing the hypothesis that $\beta_2=\beta_{2\diamond}$ can also be expressed as

$$F=\left\{\frac{RSS-USS}{k_2}\bigg/\frac{USS}{T-k}\right\}. \tag{67}$$

As a matter of interpretation, it is interesting to note that the numerator of the F statistic is also the square of the distance between $X\beta^*$, which is the estimate of the systematic component from the restricted regression, and $X\hat{\beta}$, which is its estimate from the unrestricted regression. The restricted estimate is

$$\begin{aligned} X\beta^* &= P_1(y-X_2\beta_{2\diamond})+X_2\beta_{2\diamond} \\ &= P_1y+(I-P_1)X_2\beta_{2\diamond}, \end{aligned} \tag{68}$$

and the unrestricted estimate is

$$\begin{aligned} X\hat{\beta} &= X_1\hat{\beta}_1+X_2\hat{\beta}_2 \\ &= Py. \end{aligned} \tag{69}$$

The difference between the two estimates is

$$
\begin{aligned}
X\hat{\beta} - X\beta^* &= (P - P_1)y - (I - P_1)X_2\beta_{2\diamond} \\
&= (I - P_1)(Py - X_2\beta_{2\diamond}) \\
&= (I - P_1)(X_2\hat{\beta}_2 - X_2\beta_{2\diamond}).
\end{aligned} \tag{70}
$$

Here the final identity comes from the fact that $(I - P_1)X_1\hat{\beta}_1 = 0$. It then follows from the idempotency of $(I-P_1)$ that the square of the distance between $X\beta^*$ and $X\hat{\beta}$ is

$$
(X\hat{\beta} - X\beta^*)'(X\hat{\beta} - X\beta^*) = (\hat{\beta}_2 - \beta_{2\diamond})'X_2'(I - P_1)X_2(\hat{\beta}_2 - \beta_{2\diamond}). \tag{71}
$$

The expression on the RHS repeats the expression found in (48).

Computation of the Least-squares Regression Coefficients

The methods of computing the regression coefficients which are nowadays favoured depend upon the so-called Q–R decomposition of the data matrix X. For such a matrix of full column rank, it is possible to write

$$
X = Q\begin{bmatrix} R \\ 0 \end{bmatrix}, \tag{72}
$$

where Q is an orthogonal matrix such that $Q'Q = QQ' = I$ and R is an upper (or right) triangular matrix. Amongst the methods for obtaining the decomposition are the Gram–Schmidt orthogonalisation procedure, Householder's method and Given's method. Let $Q = [Q_1, Q_2]$. Then

$$
X = [\,Q_1 \quad Q_2\,]\begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1R. \tag{73}
$$

On substituting $Q_1R = X$ into the normal equations $X'X\beta = X'y$ which determine the regression estimates, we get

$$
R'Q_1'Q_1R\beta = R'R = R'Q_1y, \tag{74}
$$

where the second equality follows in consequence of the condition that $Q_1'Q_1 = I$. Premultiplying the equations by R'^{-1} gives

$$
R\beta = Q_1y. \tag{75}
$$

Since R is an upper-triangular matrix, the equations can be solved to obtain the regression estimate $\hat{\beta}$ via a simple process of back-substitution with begins by finding the final, kth, element.

The estimate $\hat{\sigma}^2 = y'(I-P)y/(T-k)$ of the disturbance variance can also be obtained easily from the products of the Q–R decomposition. Substituting $Q_1R = X$ in the formula $P = X(X'X)^{-1}X'$ gives

$$
\begin{aligned}
P &= Q_1R(R'Q_1'Q_1R)^{-1}R'Q_1' \\
&= Q_1R(R'R)^{-1}R'Q_1' \\
&= Q_1Q_1'.
\end{aligned} \tag{76}
$$

From the fact that $QQ' = Q_1Q_1' + Q_2Q_2' = I$, it follows that $I - P = I - Q_1Q_1' = Q_2Q_2'$. Hence

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{y'(I-P)y}{T-k} \\
&= \frac{y'Q_2Q_2'y}{T-k}.
\end{aligned} \tag{77}
$$

In performing the computations, one should operate on the vector y at the same time as the matrix X is reduced to R. This will generate

$$
Q'\,[\,X \quad y\,] = \begin{bmatrix} Q_1' \\ Q_2' \end{bmatrix} [\,X \quad y\,] = \begin{bmatrix} R & Q_1'y \\ 0 & Q_2'y \end{bmatrix}. \tag{78}
$$

Thus, the components of the equations $R\beta = Q_1'y$ and $\hat{\sigma}^2 = y'Q_2Q_2'y/(T-k)$ come to hand immediately.

In practice, the transformation of X can accomplished most easily in a process of k iterations, each of which consists of premultiplying the matrix by an elementary Householder transformation which reduces all of the subdiagonal elements of a given column vector to zeros. Since each Householder transformation can be expressed as an orthonormal matrix, the product $Q' = P_kP_{k-1}\cdots P_1$ of the k transformations is itself an orthonormal matrix.

Restricted Least-Squares Regression

Sometimes, we find that there is a set of *a priori* restrictions on the elements of the vector β of the regression coefficients which can be taken into account in the process of estimation. 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, we adopt the criterion of minimising the sum of squares $(y - X\beta)'(y - X\beta)$ subject to the condition that $R\beta = r$. This leads to the Lagrangean function

$$
\begin{aligned}
L &= (y - X\beta)'(y - X\beta) + 2\lambda'(R\beta - r) \\
&= y'y - 2y'X\beta + \beta'X'X\beta + 2\lambda'R\beta - 2\lambda'r.
\end{aligned} \tag{79}
$$

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

$$2\beta' X' X - 2y' X + 2\lambda' R = 0, \tag{80}$$

whence, after transposing the expression, eliminating the factor 2 and rearranging, we have

$$X' X\beta + R'\lambda = X' y. \tag{81}$$

When these equations are compounded with the equations of the restrictions, which are supplied by the condition $\partial L/\partial\lambda = 0$, we get the following system:

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ r \end{bmatrix}. \tag{82}$$

For the system to have a unique solution, that is to say, for the existence of an estimate of β, it is not necessary that the matrix $X'X$ should be invertible—it is enough that the condition

$$\text{Rank}\begin{bmatrix} X \\ R \end{bmatrix} = k \tag{83}$$

should hold, which means that the matrix should have full column rank. The nature of this condition can be understood by considering the possibility of estimating β by applying ordinary least-squares regression to the equation

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix}\beta + \begin{bmatrix} \varepsilon \\ 0 \end{bmatrix}, \tag{84}$$

which puts the equations of the observations and the equations of the restrictions on an equal footing. It is clear that 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.

Let us simplify matters by assuming that $(X'X)^{-1}$ *does* exist. Then equation (81) gives an expression for β in the form of

$$\begin{aligned} \beta^* &= (X'X)^{-1}X'y - (X'X)^{-1}R'\lambda \\ &= \hat{\beta} - (X'X)^{-1}R'\lambda, \end{aligned} \tag{85}$$

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

$$r = R\hat{\beta} - R(X'X)^{-1}R'\lambda, \tag{86}$$

from which

$$\lambda = \{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r). \tag{87}$$

On substituting this expression back into equation (85), we get

$$\beta^* = \hat{\beta} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r). \tag{88}$$

This formula is more intelligible than it might appear to be at first, for it is simply an instance of the prediction-error algorithm whereby the estimate of β is updated in the light of the information provided by the restrictions. The error, in this instance, is the divergence between $R\hat{\beta}$ and $E(R\hat{\beta}) = r$. Also included in the formula are the terms $D(R\hat{\beta}) = \sigma^2 R(X'X)^{-1}R'$ and $C(\hat{\beta}, R\hat{\beta}) = \sigma^2(X'X)^{-1}R'$.

The sampling properties of the restricted least-squares estimator are easily established. Given that $E(\hat{\beta} - \beta) = 0$, which is to say that $\hat{\beta}$ is an unbiased estimator, it follows that $E(\beta^* - \beta) = 0$, so that β^* is also unbiased.

Next consider the expression

$$\begin{aligned} \beta^* - \beta &= [I - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R](\hat{\beta} - \beta) \\ &= (I - P_R)(\hat{\beta} - \beta), \end{aligned} \tag{89}$$

where

$$P_R = (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R. \tag{90}$$

The expression comes from taking β from both sides of (88) and from recognising that $R\hat{\beta} - r = R(\hat{\beta} - \beta)$. We may observe that P_R is an idempotent matrix which is subject to the conditions that

$$P_R = P_R^2, \quad P_R(I - P_R) = 0 \quad \text{and} \quad P_R'X'X(I - P_R) = 0. \tag{91}$$

From equation (89), we deduce that

$$\begin{aligned} D(\beta^*) &= (I - P_R)E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\}(I - P_R) \\ &= \sigma^2(I - P_R)(X'X)^{-1}(I - P_R) \\ &= \sigma^2[(X'X)^{-1} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}]. \end{aligned} \tag{92}$$

Restricted Least Squares and the Decomposition of a Chi-Square

Consider the identity

$$\begin{aligned} y - X\beta &= (y - X\hat{\beta}) + (X\hat{\beta} - X\beta) \\ &= (y - X\hat{\beta}) + (X\hat{\beta} - X\beta^*) + (X\beta^* - X\beta). \end{aligned} \tag{93}$$

Here, there are

$$
(94) \qquad \begin{aligned} X\beta^* - X\beta &= X(I - P_R)(\hat{\beta} - \beta), \quad \text{and} \\ X\hat{\beta} - X\beta^* &= XP_R(\hat{\beta} - \beta). \end{aligned}
$$

The first of these comes directly from (89), whereafter the second is also implied. On substituting for $\hat{\beta} - \beta = (X'X)^{-1}X\varepsilon$, we get

$$
(95) \qquad \begin{aligned} X\beta^* - X\beta &= X(I - P_R)(X'X)^{-1}X\varepsilon = (P - P_Q)\varepsilon, \quad \text{and} \\ X\hat{\beta} - X\beta^* &= XP_R(X'X)^{-1}X\varepsilon = P_Q\varepsilon. \end{aligned}
$$

Here, we have defined

$$
(96) \qquad \begin{aligned} P_Q &= XP_R(X'X)^{-1}X' \\ &= X(X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}X'. \end{aligned}
$$

which is a symmetric idempotent matrix fulfilling the conditions that

$$
(97) \qquad P_Q = P_Q^2 = P_Q' \quad \text{and} \quad PP_Q = P_QP = P_Q.
$$

The decompositions of (93) can be represented in terms of the various symmetric idempotent projection operators defined above. Thus,

$$
(98) \qquad \varepsilon = (I - P)\varepsilon + (P - P_Q)\varepsilon + P_Q\varepsilon;
$$

and, since the symmetric idempotent matrices $I - P$, $P - P_Q$ and P_Q are mutually orthogonal, It follow that

$$
(99) \qquad \varepsilon'\varepsilon = \varepsilon'(I - P)\varepsilon + \varepsilon'(P - P_Q)\varepsilon + \varepsilon'P_Q\varepsilon.
$$

Moreover, if $y - X\beta = \varepsilon \sim N(0, \sigma^2 I)$, then, according to Cochrane's theorem, we should have

$$
(100) \qquad \begin{aligned} &\text{(a)} \qquad \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) \sim \sigma^2\chi^2(T) \\ &\text{(b)} \qquad \varepsilon'(I - P)\varepsilon = (y - X\hat{\beta})'(y - X\hat{\beta}) \sim \sigma^2\chi^2(T - k) \\ &\text{(c)} \qquad \varepsilon'(P - P_Q)\varepsilon = (X\beta^* - X\beta)'(X\beta^* - X\beta) \sim \sigma^2\chi^2(k - j) \\ &\text{(d)} \qquad \varepsilon'P_Q\varepsilon = (X\hat{\beta} - X\beta^*)'(X\hat{\beta} - X\beta^*) \sim \sigma^2\chi^2(j) \end{aligned}
$$

Here (b), (c) and (d) represent statistically independent random variables whose sum is the random variable of (a).

From the results under (b) and (d), one can derive estimators of the disturbance variance $V(\varepsilon_t) = \sigma^2$. The estimator $\hat{\sigma}^2$ of (19) comes directly from (b). For an alternative estimator, we may consider the identity

$$\begin{aligned} y - X\beta^* &= (y - X\hat{\beta}) + (X\hat{\beta} - X\beta^*) \\ &= (I - P)\varepsilon + P_Q\varepsilon. \end{aligned} \tag{101}$$

On the LHS is the sum of two mutually orthogonal vector components whose sums of squares give rise to statistically independent chi-square variates. The sum of the chi-squares is itself a chi-square:

$$\begin{aligned} (y - X\beta^*)'(y - X\beta^*) &= (y - X\hat{\beta})'(y - X\hat{\beta}) + (X\hat{\beta} - X\beta^*)'(X\hat{\beta} - X\beta^*) \\ &= \varepsilon'(I - P)\varepsilon + \varepsilon P_Q\varepsilon \sim \sigma^2\chi^2(T - K + j). \end{aligned} \tag{102}$$

The expected value of this quadratic is $\sigma^2(T - K + j)$, and it follows that

$$\sigma^{*2} = \frac{1}{T - k + j}(y - X\beta^*)'(y - X\beta^*), \tag{103}$$

is an unbiased estimator of the variance.

This inference follows from the fact that the expected value of a chi-square variate of r degrees of freedom is r. A demonstration of the unbiasedness of the estimator is available which makes no reference to the functional form of its distribution and which is similar to the demonstration of the unbiasedness of $\hat{\sigma}^2$. Under the assumption of a chi-square distribution, the variances of the two estimators are $V(\hat{\sigma}^2) = 2\sigma^4/(T - k)$ and $V(\hat{\sigma}^{*2}) = 2\sigma^4/(T - k + j)$; and so, provided that the restrictions are valid, σ^{*2} is the more efficient estimator.

Testing the Linear Restrictions

Given that

$$\begin{aligned} &\sigma^{-2}(\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*) \sim \chi^2(j) \quad \text{and} \\ &\sigma^{-2}(y - X\hat{\beta})'(y - X\hat{\beta}) \sim \chi^2(T - k) \end{aligned} \tag{104}$$

are mutually independent chi-square variates, it follows that

$$\begin{aligned} F &= \left\{ \frac{(\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*)}{j} \bigg/ \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T - k} \right\} \\ &= \frac{1}{\hat{\sigma}^2 j}(\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*) \sim F(j, T - k) \end{aligned} \tag{105}$$

has a F distribution of j and $T-k$ degrees of freedom. This statistic may be used to test the validity of the restrictions which are incorporated in the estimate β^*. However, the test does not depend upon finding the value of the restricted estimate. From equation (88), it follows that

$$(106)\qquad \frac{(X\hat{\beta}-X\beta^*)'(X\hat{\beta}-X\beta^*)}{\hat{\sigma}^2 j}=\frac{(R\hat{\beta}-r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta}-r)}{\hat{\sigma}^2 j},$$

The form on the RHS can also be derived in straightforward manner by considering the distribution of the unrestricted estimator. From the fact that $\hat{\beta}\sim N\{\beta,\sigma^2(X'X)^{-1}\}$, it follows that

$$(107)\qquad R\hat{\beta}\sim N\{R\beta=r,\sigma^2R(X'X)^{-1}R'\}.$$

We infer immediately that

$$(108)\qquad \frac{(R\hat{\beta}-r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta}-r)}{\sigma^2}\sim\chi^2(j).$$

Yet another way of expressing the numerator is to use the identity

$$(109)\quad (X\hat{\beta}-X\beta^*)'(X\hat{\beta}-X\beta^*)=(y-X\beta^*)'(y-X\beta^*)-(y-X\hat{\beta})'(y-X\hat{\beta}).$$

To establish this identity, it is sufficient to observe that $y-X\beta^*=(y-X\hat{\beta})+(X\hat{\beta}+X\beta^*)$ and to show that $(y-X\hat{\beta})\perp(X\hat{\beta}+X\beta^*)$. Then the result follows from Pythagoras' theorem. The condition of orthogonality follows from writing $y-X\hat{\beta}=(I-P)\varepsilon$ and $X\hat{\beta}+X\beta^*=P_Q\varepsilon$ and from noting that, according to (97), $P_Q'(I-P)=P_Q-P_QP=0$.

The quantity $(y-X\beta^*)'(y-X\beta^*)$ on the RHS of (109) is the restricted sum of squares denoted by RSS, whilst the quantity $(y-X\hat{\beta})'(y-X\hat{\beta})$ is the unrestricted sum of squares denoted by USS. Thus, the F statistic of (105) can be expressed as

$$(110)\qquad F=\left\{\frac{RSS-USS}{j}\bigg/\frac{USS}{T-k}\right\},$$

which is comparable to the expression under (67).

Example. A specialisation of the statistic on the RHS of (106) 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_2 assumes the value of $\beta_{2\diamond}$ can be expressed via the equation

$$(111)\qquad [0,I_{k_2}]\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}=\beta_{2\diamond}.$$

This can be construed as a case of the equation $R\beta = r$ where $R = [0, I_{k_2}]$ and $r = \beta_{2\diamond}$.

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

$$
\begin{aligned}
(X'X)^{-1} &= \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \{X_1'(I-P_2)X_1\}^{-1} & -\{X_1'(I-P_2)X_1\}^{-1}X_1'X_2(X_2'X_2)^{-1} \\ -\{X_2'(I-P_1)X_2\}^{-1}X_2'X_1(X_1'X_1)^{-1} & \{X_2'(I-P_1)X_2\}^{-1} \end{bmatrix},
\end{aligned} \tag{112}
$$

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

$$
R(X'X)^{-1}R' = \{X_2'(I-P_1)X_2\}^{-1} \tag{113}
$$

It follows in a straightforward manner that the specialised form of the F statistic of (106) is

$$
\begin{aligned}
F &= \left\{ \frac{(\hat{\beta}_2 - \beta_{2\diamond})'X_1'(I-P_2)X_1(\hat{\beta}_2 - \beta_{2\diamond})}{k_2} \Big/ \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T-k} \right\} \\
&= \frac{(\hat{\beta}_2 - \beta_{2\diamond})'X_2'(I-P_1)X_2'(\hat{\beta}_2 - \beta_{2\diamond})}{\hat{\sigma}^2 k_2} \sim F(k_2, T-k).
\end{aligned} \tag{114}
$$

This is a test statistic that has been presented previously under (48).

Example. A example of the testing of linear restrictions is provided by the so-called Chow test which is aimed at uncovering structural breaks which involve abrupt changes in the vector of regression coefficients. The unrestricted model which accommodates a structural break can be written as

$$
\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}. \tag{111}
$$

The restricted model, which excludes the possibility of a break, is written as

$$
\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}. \tag{112}
$$

Within the context of equation (111), the restriction which is to be tested takes the form of

$$
[\,I \quad -I\,] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = 0. \tag{113}
$$

Computation of the Restricted Least-squares Estimates

The formula of (88) for the restricted least-squares estimate has an appearance which suggests that process of computing the coefficients is liable to be a laborious one. In practice, the task can be greatly simplified.

The procedure which we shall propose uses the equation of the restrictions to reparametrise the regression equations in a way that reduces the dimension of the coefficient vector which has to be estimated. From this vector of reduced dimension, the restricted estimate of the ordinary coefficient vector can be recovered.

The first step is to transform the restriction $R\beta = r$ to a more convenient form which depends upon the Q–R decomposition of R'. This is

$$R' = C\begin{bmatrix} U \\ 0 \end{bmatrix} = [\,C_1, C_2\,]\begin{bmatrix} U \\ 0 \end{bmatrix} = C_1 U, \tag{114}$$

where $C = [C_1, C_2]$ is an orthonormal matrix such that $C'C = CC' = I$, and U is an upper triangular matrix. Then $R\beta = U_1' C_1' \beta = r$, or , equivalently, $C_1'\beta = (U_1')^{-1} r = h$. Thus, the restriction can be written alternatively as

$$C_1'\beta = h. \tag{115}$$

Now, the condition that $C'C = I$ implies that $C_1'C_1 = I$ and that $C_1'C_2 = 0$; and thus it follows that a solution of the equations (115) must take the form of

$$\beta = C_2\gamma + C_1 h, \tag{116}$$

where γ is an undetermined vector of order $k - j$. It follows that

$$\begin{aligned} y &= X\beta + \varepsilon \\ &= XC_2\delta + XC_1 h + \varepsilon. \end{aligned} \tag{117}$$

This becomes

$$q = X\gamma + \varepsilon, \quad \text{where} \quad q = y - XC_1 h \quad \text{and} \quad Z = XC_2. \tag{118}$$

The ordinary least-squares estimator of γ can be found by the methods, already described, which are appropriate to the unrestricted regression model. The restricted estimate of β can be found by substituting the estimate $\hat{\gamma}$ into equation (116).