LECTURE 5

Hypothesis Testing in the Classical Regression Model

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:

(1) 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,

(2) 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,

(3) 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,

(4) 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,

(5) 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. The oridinary least squares estimate $\beta = (X'X)^{-1}X'y$ of the parameter vector β in the regression model $(y; X\beta, \sigma^2 I)$ is a linear function of y, which has an expected value of $E(\hat{\beta}) = \beta$ and a dispersion matrix of $D(\hat{\beta}) = \sigma^2(X'X)^{-1}$, Thus, it follows that, if $y \sim N(X\beta, \sigma^2 I)$ is normally distributed, then

$$\hat{\beta} \sim N_k\{\beta, \sigma^2(X'X)^{-1}\}. \tag{6}$$

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

$$\hat{\beta}_1 \sim N_{k_1}\big(\beta_1, \sigma^2\{X_1'(I - P_2)X_1\}^{-1}\big), \tag{7}$$

$$\hat{\beta}_2 \sim N_{k_2}\big(\beta_2, \sigma^2\{X_2'(I - P_1)X_2\}^{-1}\big). \tag{8}$$

From the results under (2) to (6), it follows that

$$\sigma^{-2}(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \sim \chi^2(k). \tag{9}$$

Similarly, it follows from (7) and (8) that

$$\sigma^{-2}(\hat{\beta}_1 - \beta_1)'X_1'(I - P_2)X_1(\hat{\beta}_1 - \beta_1) \sim \chi^2(k_1), \tag{10}$$

$$\sigma^{-2}(\hat{\beta}_2 - \beta_2)'X_2'(I - P_1)X_2(\hat{\beta}_2 - \beta_2) \sim \chi^2(k_2). \tag{11}$$

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 residual

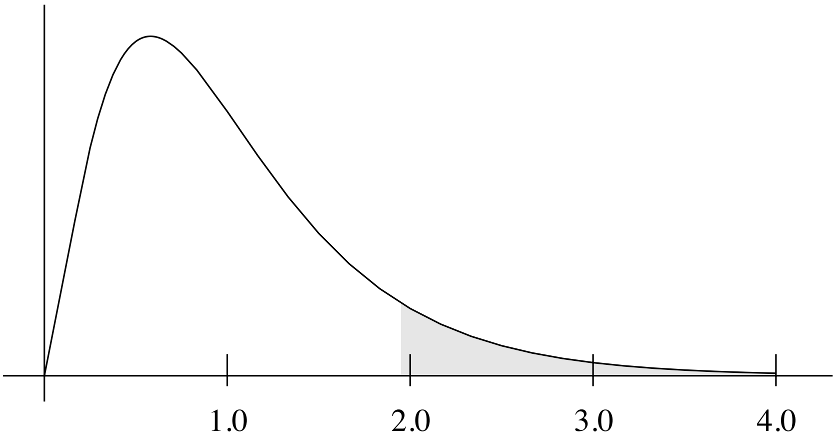

Figure 1. The critical region, at the 10% significance level, of an $F(5, 60)$ statistic.

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 $C'X = 0$. Since $C'C = I_{T-k}$, 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

$$\sigma^{-2} y'CC'y = \sigma^{-2}(y - X\hat{\beta})'(y - X\hat{\beta}) \sim \chi^2(T - k). \tag{12}$$

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

$$\begin{aligned} &\sigma^{-2}(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \sim \chi^2(k) \quad \text{and} \\ &\sigma^{-2}(y - X\hat{\beta})'(y - X\hat{\beta}) \sim \chi^2(T - k) \end{aligned} \tag{13}$$

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

$$\begin{aligned} F &= \left\{ \frac{(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)}{k} \Bigg/ \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T - k} \right\} \\ &= \frac{1}{\hat{\sigma}^2 k}(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \sim F(k, T - k). \end{aligned} \tag{14}$$

To test an hypothesis specifying that $\beta = \beta_\diamond$, the hypothesised parameter vector can be inserted in the above statistic and the resulting value can be compared 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 $X\hat{\beta}$ that 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 that 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

$$F = \frac{1}{\hat{\sigma}^2 k_2}(\hat{\beta}_2 - \beta_{2\diamond})' X_2'(I - P_1)X_2(\hat{\beta}_2 - \beta_{2\diamond}). \tag{15}$$

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 (15), a statistic may be derived in the form of

$$F = \frac{(\hat{\beta}_i - \beta_{i\diamond})^2}{\hat{\sigma}^2 w_{ii}}, \tag{16}$$

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

$$t = \frac{\hat{\beta}_i - \beta_{i\diamond}}{\sqrt{(\hat{\sigma}^2 w_{ii})}} \tag{17}$$

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

$$\begin{aligned} y &= X\hat{\beta} + (y - X\hat{\beta}) \quad \text{and} \\ y - X\beta &= (X\hat{\beta} - X\beta) + (y - X\hat{\beta}), \end{aligned} \tag{18}$$

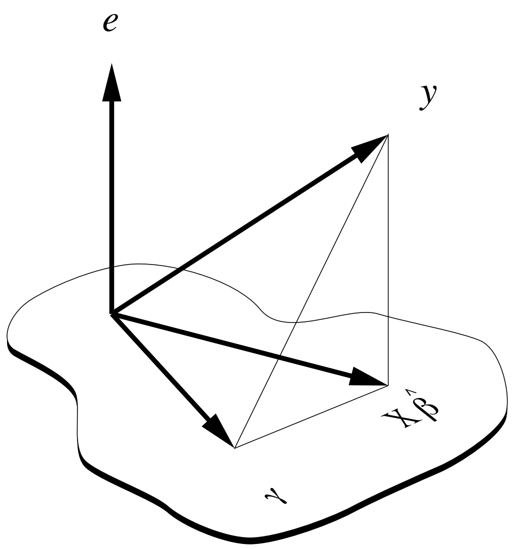

Figure 2. The vector $Py = X\hat{\beta}$ is formed by the orthogonal projection of the vector y onto the subspace spanned by the columns of the matrix X.

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,

$$(19)\qquad \begin{aligned} y &= Py + (I - P)y \qquad \text{and} \\ \varepsilon &= P\varepsilon + (I - P)\varepsilon. \end{aligned}$$

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

$$(20)\qquad \begin{aligned} \varepsilon'\varepsilon &= \varepsilon' P\varepsilon + \varepsilon'(I - P)\varepsilon \qquad \text{or, equivalently,} \\ \varepsilon'\varepsilon &= (X\hat{\beta} - X\beta)'(X\hat{\beta} - X\beta) + (y - X\hat{\beta})'(y - X\hat{\beta}). \end{aligned}$$

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 2, 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, the value of $X\beta_\diamond$ is compared 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 that separates them. This would be given by

$$\varepsilon' P\varepsilon = (X\hat{\beta} - X\beta_\diamond)'(X\hat{\beta} - X\beta_\diamond). \tag{21}$$

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

$$\begin{aligned} \hat{\sigma}^2 &= \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T - k} \\ &= \frac{\varepsilon'(I - P)\varepsilon}{T - k}. \end{aligned} \tag{22}$$

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

$$\begin{aligned} &\text{(a)} \quad \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) \sim \sigma^2\chi^2(T), \\ &\text{(b)} \quad \varepsilon' P\varepsilon = (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \sim \sigma^2\chi^2(k), \\ &\text{(c)} \quad \varepsilon'(I - P)\varepsilon = (y - X\hat{\beta})'(y - X\hat{\beta}) \sim \sigma^2\chi^2(T - k), \end{aligned} \tag{23}$$

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 (14) which is

$$F = \left\{ \frac{\varepsilon' P\varepsilon}{k} \bigg/ \frac{\varepsilon'(I - P)\varepsilon}{T - k} \right\} \sim F(k, T - k). \tag{24}$$

This result depends upon Cochrane's Theorem concerning the decomposition of a chi-square random variate. The following is a statement of the theorem which is attuned to the present requirements:

(25) Let $\varepsilon \sim N(0, \sigma^2 I_T)$ be a random vector of T independently and identically distributed elements. Also, let $P = X(X'X)^{-1}X'$ be a symmetric idempotent matrix, such that $P = P' = P^2$, which is constructed from a matrix X of order $T \times k$ with $\text{Rank}(X) = k$. Then

$$\frac{\varepsilon' P \varepsilon}{\sigma^2} + \frac{\varepsilon'(I - P)\varepsilon}{\sigma^2} = \frac{\varepsilon'\varepsilon}{\sigma^2} \sim \chi^2(T),$$

which is a chi-square variate of T degrees of freedom, represents the sum of two independent chi-square variates $\varepsilon' P \varepsilon / \sigma^2 \sim \chi^2(k)$ and $\varepsilon'(I - P)\varepsilon/\sigma^2 \sim \chi^2(T - k)$ of k and $T - k$ degrees of freedom respectively.

To prove this result, we begin by finding an alternative expression for the projector $P = X(X'X)^{-1}X'$. First, consider the fact that $X'X$ is a symmetric positive-definite matrix. It follows that there exists a matrix transformation T such that $T(X'X)T' = I$ and $T'T = (X'X)^{-1}$. Therefore, $P = XT'TX' = C_1C_1'$, where $C_1 = XT'$ is a $T \times k$ matrix comprising k orthonormal vectors such that $C_1'C_1 = I_k$ is the identity matrix of order k.

Now define C_2 to be a complementary matrix of $T-k$ orthonormal vectors. Then, $C = [C_1, C_2]$ is an orthonormal matrix of order T such that

$$\begin{aligned} CC' &= C_1C_1' + C_2C_2' = I_T \qquad \text{and} \\ C'C &= \begin{bmatrix} C_1'C_1 & C_1'C_2 \\ C_2'C_1 & C_2'C_2 \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ 0 & I_{T-k} \end{bmatrix}. \end{aligned} \tag{26}$$

The first of these results allows us to set $I - P = I - C_1C_1' = C_2C_2'$. Now, if $\varepsilon \sim N(0, \sigma^2 I_T)$ and if C is an orthonormal matrix such that $C'C = I_T$, then it follows that $C'\varepsilon \sim N(0, \sigma^2 I_T)$. In effect, if ε is a normally distributed random vector with a density function which is centred on zero and which has spherical contours, and if C is the matrix of a rotation, then nothing is altered by applying the rotation to the random vector. On partitioning $C'\varepsilon$, we find that

$$\begin{bmatrix} C_1'\varepsilon \\ C_2'\varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 I_k & 0 \\ 0 & \sigma^2 I_{T-k} \end{bmatrix} \right), \tag{27}$$

which is to say that $C_1'\varepsilon \sim N(0, \sigma^2 I_k)$ and $C_2'\varepsilon \sim N(0, \sigma^2 I_{T-k})$ are independently distributed normal vectors. It follows that

$$\begin{aligned} \frac{\varepsilon' C_1 C_1' \varepsilon}{\sigma^2} &= \frac{\varepsilon' P \varepsilon}{\sigma^2} \sim \chi^2(k) \quad \text{and} \\ \frac{\varepsilon' C_2 C_2' \varepsilon}{\sigma^2} &= \frac{\varepsilon'(I - P)\varepsilon}{\sigma^2} \sim \chi^2(T - k) \end{aligned} \tag{28}$$

are independent chi-square variates. Since $C_1C_1' + C_2C_2' = I_T$, the sum of these two variates is

$$(29)\qquad \frac{\varepsilon' C_1 C_1' \varepsilon}{\sigma^2} + \frac{\varepsilon' C_2 C_2' \varepsilon}{\sigma^2} = \frac{\varepsilon'\varepsilon}{\sigma^2} \sim \chi^2(T);$$

and thus the theorem is proved.

The statistic under (14) can now be expressed in the form of

$$(30)\qquad F = \left\{ \frac{\varepsilon' P \varepsilon}{k} \Big/ \frac{\varepsilon'(I-P)\varepsilon}{T-k} \right\}.$$

This is manifestly the ratio of two chi-square variates divided by their respective degrees of freedom; and so it has an F distribution with these degrees of freedom. This result provides the means for testing the hypothesis concerning the parameter vector β.

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

$$(31)\qquad R\beta = r,$$

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. Given that the ordinary least-squares estimator of β is a normally distributed vector $\hat\beta \sim N\{\beta, \sigma^2(X'X)^{-1}\}$, it follows that

$$(32)\qquad R\hat\beta \sim N\big\{R\beta = r, \sigma^2 R(X'X)^{-1}R'\big\};$$

and, from this, it can be inferred immediately that

$$(33)\qquad \frac{(R\hat\beta - r)'\big\{R(X'X)^{-1}R'\big\}^{-1}(R\hat\beta - r)}{\sigma^2} \sim \chi^2(j).$$

It has already established been established that

$$(34)\qquad \frac{(T-k)\hat\sigma^2}{\sigma^2} = \frac{(y - X\hat\beta)'(y - X\hat\beta)}{\sigma^2} \sim \chi^2(T-k)$$

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

$$(35)\qquad \frac{(\hat\beta - \beta)'X'X(\hat\beta - \beta)}{\sigma^2} \sim \chi^2(k)$$

derived from the estimator of the regression parameters. The variate of (33) must also be independent of the chi-square of (34); and it is straightforward to deduce that

$$
\begin{aligned}
F &= \left\{ \frac{(R\hat{\beta}-r)'\left\{R(X'X)^{-1}R'\right\}^{-1}(R\hat{\beta}-r)}{j} \bigg/ \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{T-k} \right\} \\
&= \frac{(R\hat{\beta}-r)'\left\{R(X'X)^{-1}R'\right\}^{-1}(R\hat{\beta}-r)}{\hat{\sigma}^2 j} \sim F(j, T-k),
\end{aligned} \tag{36}
$$

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 (36) 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 $\beta_1^\diamond$ can be expressed via the equation

$$
[I_{k_1}, 0] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \beta_1^\diamond. \tag{37}
$$

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

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

$$
\begin{aligned}
(X'X)^{-1} &= \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \{X_1'(I-P_2)X_1\}^{-1} & -\{X_1'(I-P_2)X_1\}^{-1}X_1'X_2(X_2'X_2)^{-1} \\ -\{X_2'(I-P_1)X_2\}^{-1}X_2'X_1(X_1'X_1)^{-1} & \{X_2'(I-P_1)X_2\}^{-1} \end{bmatrix},
\end{aligned} \tag{38}
$$

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

$$
R(X'X)^{-1}R' = \left\{X_1'(I-P_2)X_1\right\}^{-1}. \tag{39}
$$

It follows, in a straightforward manner, that the specialised form of the F statistic of (36) is

$$
\begin{aligned}
F &= \left\{ \frac{(\hat{\beta}_1-\beta_1^\diamond)'\left\{X_1'(I-P_2)X_1\right\}(\hat{\beta}_1-\beta_1^\diamond)}{k_1} \bigg/ \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{T-k} \right\} \\
&= \frac{(\hat{\beta}_1-\beta_1^\diamond)'\left\{X_1'(I-P_2)X_1'\right\}(\hat{\beta}_1-\beta_1^\diamond)}{\hat{\sigma}^2 k_1} \sim F(k_1, T-k).
\end{aligned} \tag{40}
$$

Finally, for the jth element of $\hat{\beta}$, there is

$$(41)\qquad \begin{aligned}&(\hat{\beta}_j - \beta_j)^2/\sigma^2 w_{jj} \sim F(1, T-k) \quad \text{or, equivalently,}\\ &(\hat{\beta}_j - \beta_j)\sqrt{\sigma^2 w_{jj}} \sim t(T-k),\end{aligned}$$

where w_{jj} is the jth diagonal element of $(X'X)^{-1}$ and $t(T-k)$ denotes the t distribution of $T-k$ degrees of freedom.

An Alternative Formulation of the F statistic

An alternative way of forming the F statistic uses the products of two separate regressions. Consider the formula for the restricted least-squares estimator that has been given under (2.76):

$$(42)\qquad \beta^* = \hat{\beta} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r).$$

From this, the following expression for the residual sum of squares of the restricted regression is is derived:

$$(43)\qquad y - X\beta^* = (y - X\hat{\beta}) + X(X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r).$$

The two terms on the RHS are mutually orthogonal on account of the defining condition of an ordinary least-squares regression, which is that $(y - X\hat{\beta})'X = 0$. Therefore, the residual sum of squares of the restricted regression is

$$(44)\qquad \begin{aligned}(y - X\beta^*)'(y - X\beta^*) = (y - X\hat{\beta})'(y - X\hat{\beta}) &+ \\ (R\hat{\beta} - r)'\big\{R(X'X)^{-1}R'\big\}^{-1}&(R\hat{\beta} - r).\end{aligned}$$

This equation can be rewritten as

$$(45)\qquad RSS - USS = (R\hat{\beta} - r)'\big\{R(X'X)^{-1}R'\big\}^{-1}(R\hat{\beta} - r),$$

where RSS denotes the restricted sum of squares an USS denotes the unrestricted sum of squares. It follows that the test statistic of (36) can be written as

$$(46)\qquad F = \left\{\frac{RSS - USS}{j} \bigg/ \frac{USS}{T-k}\right\}.$$

This formulation can be used, for example, in testing the restriction that $\beta_1 = 0$ in the partitioned model $N(y; X_1\beta_1 + X_2\beta_2, \sigma^2 I)$. Then, in terms of equation (37), there is $R = [I_{k_1}, 0]$ and there is $r = \beta_1^\diamond = 0$, which gives

$$(47)\qquad \begin{aligned}RSS - USS &= \hat{\beta}_1' X_1'(I - P_2)X_1\hat{\beta}_1\\ &= y'(I - P_2)X_1\{X_1'(I - P_2)X_1\}^{-1}X_1'(I - P_2)y.\end{aligned}$$

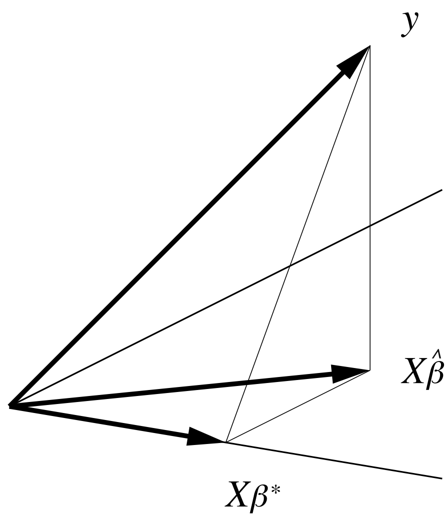

Figure 3. The test of the hypothesis entailed by the resticted model is based on a measure of the proximity of the restricted estimate $X\beta^*$, and the unrestricted estimate $X\hat{\beta}$. The USS is the squared distance $\|y-X\hat{\beta}\|^2$. The RSS is the squared distance $\|y - X\beta^*\|^2$.

On the other hand, there is

$$RSS - USS = y'(I - P_2)y - y'(I - P)y = y'(P - P_2)y, \tag{48}$$

Since the two expressions must be identical for all values of y, the comparison of (36) and (37) is sufficient to establish the following identity:

$$(I - P_2)X_1\{X_1'(I - P_2)X_1\}^{-1}X_1'(I - P_2) = P - P_2. \tag{49}$$

The geometric interpretation of the alternative formulation of the test statistic is straightforward. It can be understood, in reference to Figure 3, that the square of the distance between the restricted estimate $X\beta^*$ and the unrestricted estimate $X\hat{\beta}$, denoted by $\|X\hat{\beta} - X\beta^*\|^2$, which is the basis of the original formulation of the test statistic, is equal to the restricted sum of squares $\|y - X\beta^*\|^2$ less the unrestricted sum of squares $\|y - X\hat{\beta}\|^2$. The latter is the basis of the alternative formulation.

The Partitioned Inverse and Associated Identities

The first objective is to derive the formula for the partitioned inverse of $X'X$ that has been given in equation (38). Write

$$\begin{bmatrix} A & B' \\ B & C \end{bmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \tag{50}$$

and consider, the equation

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}\begin{bmatrix} A & B' \\ B & C \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \tag{51}$$

From this system, the following two equations can be extracted:

$$X_1'X_1A + X_1'X_2B = I, \tag{52}$$

$$X_2'X_1A + X_2'X_2B = 0. \tag{53}$$

To isolate A, equation (53) is premultiplied by $X_1'X_2(X_2'X_2)^{-1}$ to give

$$X_1'X_2(X_2'X_2)^{-1}X_2'X_1A + X_1'X_2B = 0, \tag{54}$$

and this is taken from (52) to give

$$\{X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1\}A = I \tag{55}$$

whence

$$A = \{X_1'(I - P_2)X_1\}^{-1} \quad \text{with} \quad P_2 = X_2(X_2'X_2)^{-1}X_2'. \tag{56}$$

An argument of symmetry will serve to show that

$$C = \{X_2'(I - P_1)X_2\}^{-1} \quad \text{with} \quad P_1 = X_1(X_1'X_1)^{-1}X_1'. \tag{57}$$

To find B, and therefore B', the expression for A from (56) is substituted into (53). This gives

$$X_2'X_1\{X_1'(I - P_2)X_1\}^{-1} + X_2'X_2B = 0, \tag{58}$$

whence

$$B' = -\{X_1'(I - P_2)X_1\}^{-1}X_1'X_2(X_2'X_2)^{-1} \tag{59}$$

The matrix B is the transpose of this, but an argument of symmetry will serve to show that this is also given by the expression

$$B = -\{X_2'(I - P_1)X_2\}^{-1}X_2'X_1(X_1'X_1)^{-1} \tag{60}$$

When the expression for A, B, B' and C are put in place, the result is

$$\begin{aligned} &\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} A & B' \\ B & C \end{bmatrix} \\ &= \begin{bmatrix} \{X_1'(I - P_2)X_1\}^{-1} & -\{X_1'(I - P_2)X_1\}^{-1}X_1'X_2(X_2'X_2)^{-1} \\ -\{X_2'(I - P_1)X_2\}^{-1}X_2'X_1(X_1'X_1)^{-1} & \{X_2'(I - P_1)X_2\}^{-1} \end{bmatrix}, \end{aligned} \tag{61}$$

Next consider

$$
\begin{aligned}
X(X'X)^{-1}X' &= [\,X_1 \quad X_2\,]\begin{bmatrix} A & B' \\ B & C \end{bmatrix}\begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \\
&= \{X_1AX_1' + X_1BX_2'\} + \{X_2CX_2' + X_2B'X_1'\}
\end{aligned} \tag{62}
$$

Substituting the expressions for A, B, C and B' shows that

$$
\begin{aligned}
X(X'X)^{-1}X' =& X_1\{X_1'(I-P_2)X_1\}^{-1}X_1'(I-P_2) \\
&+ X_2\{X_2'(I-P_1)X_2\}^{-1}X_2'(I-P_1),
\end{aligned} \tag{63}
$$

which can be written as

$$P = P_{1/2} + P_{2/1} \tag{64}$$

with

$$P_{1/2} = X_1\{X_1'(I-P_2)X_1\}^{-1}X_1'(I-P_2), \tag{65}$$

$$P_{2/1} = X_2\{X_2'(I-P_1)X_2\}^{-1}X_2'(I-P_1). \tag{66}$$

Next, observe that there are

$$P_{1/2}P_1 = P_1, \qquad P_{1/2}P_2 = 0, \tag{67}$$

$$P_{2/1}P_2 = P_2, \qquad P_{2/1}P_1 = 0. \tag{68}$$

It follows that

$$PP_1 = (P_{1/2} + P_{2/1})P_1 = P_1 = P_1P, \tag{69}$$

where the final equality follows in consequence of the symmetry of P_1 and P. It also follows, by an argument of symmetry, that

$$P(I-P_1) = P - P_1 = (I-P_1)P \tag{70}$$

Therefore, since $(I-P_1)P = (I-P_1)P_{2/1}$, there is

$$P - P_1 = (I-P_1)X_2\{X_2'(I-P_1)X_2\}^{-1}X_2(I-P_1). \tag{71}$$

By interchanging the two subscripts, the identity of (49) is derived.