HYPOTHESIS TESTS FOR THE CLASSICAL LINEAR MODEL

The Normal Distribution and the Sampling Distributions

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 standard normal vector $z \sim N(0, I)$ has $E(x) = 0$ and $D(x) = I$. 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 $T(x - \mu) \sim N(0, I)$.

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

$$(x - \mu)'\Sigma^{-1}(x - \mu) \sim \chi^2(n).$$

(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. Thus,

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

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

The OLS estimate $\beta = (X'X)^{-1}X'y$ of β in the model $(y; X\beta, \sigma^2 I)$ has $E(\hat{\beta}) = \beta$ and $D(\hat{\beta}) = \sigma^2(X'X)^{-1}$, Thus, if $y \sim N(X\beta, \sigma^2 I)$, 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

$$\hat{\beta}_1 \sim N_{k_1}\left(\beta_1, \sigma^2\{X_1'(I - P_2)X_1\}^{-1}\right), \qquad P_2 = X_2(X_2'X_2)^{-1}X_2, \tag{7}$$

$$\hat{\beta}_2 \sim N_{k_2}\left(\beta_2, \sigma^2\{X_2'(I - P_1)X_2\}^{-1}\right), \qquad P_1 = X_1(X_1'X_1)^{-1}X_1. \tag{8}$$

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 residual vector $e = y - X\hat{\beta}$ has $E(e) = 0$ and $D(e) = \sigma^2(I - P)$ which is singular. Here, $P = X(X'X)^{-1}X$ and $I - P = C_1C_1'$, where C_1 is a $T \times (T - k)$ matrix of $T - k$ orthonormal columns, which are orthogonal to the columns of X such that $C_1'X = 0$.

Since $C_1'C_1 = I_{T-k}$, it follows that, if $y \sim N_T(X\beta, \sigma^2 I)$, then $C_1'y \sim N_{T-k}(0, \sigma^2 I)$. Hence

$$\sigma^{-2}y'C_1C_1'y = \sigma^{-2}(y - X\hat{\beta})'(y - X\hat{\beta}) \sim \chi^2(T - k). \tag{12}$$

Since they have a zero-valued covariance matrix, $X\hat{\beta} = Py$ and $y - X\hat{\beta} = (I - P)y$ are statistically independent. 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 that $\beta = \beta_\diamond$, the hypothesised value $\beta_\diamond$ 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 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.

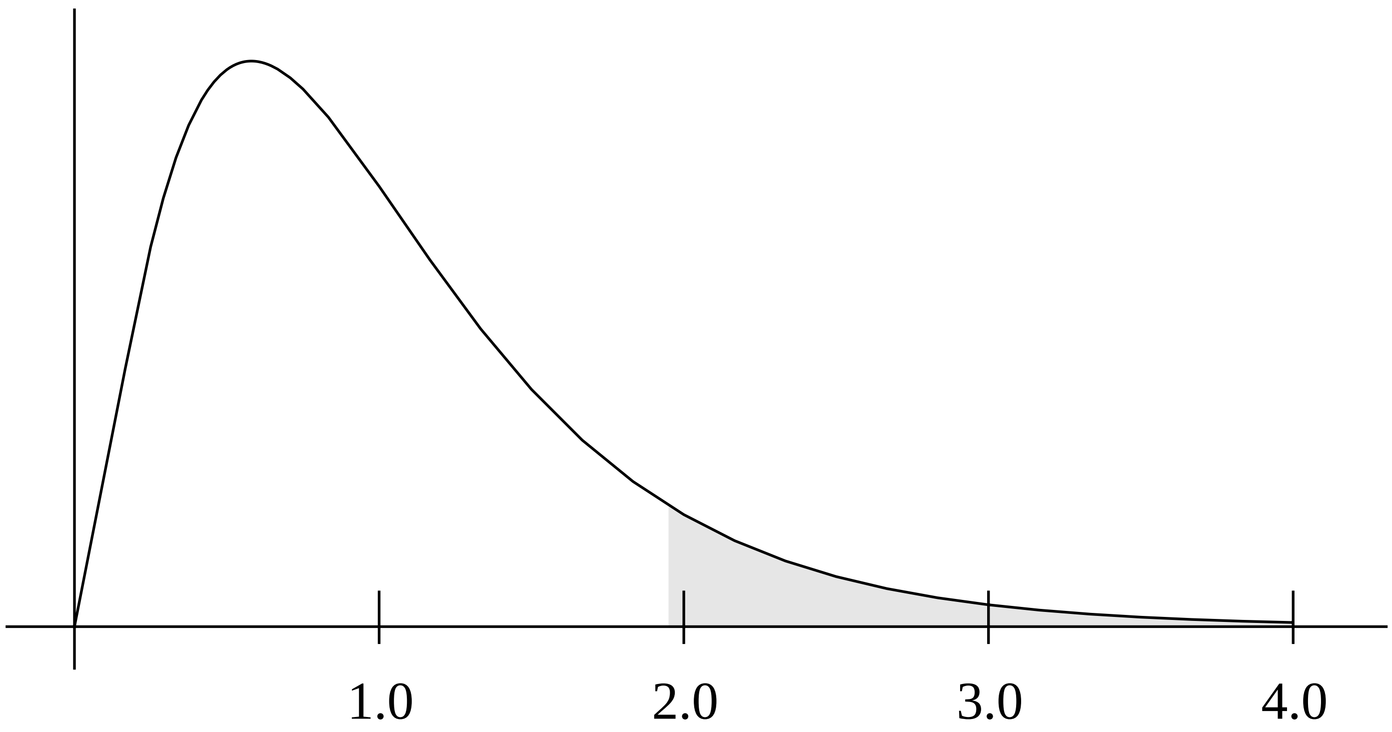

Figure 1. The critical region, at the 10% significance level, of an $F(5, 60)$ statistic.

To test an hypothesis that $\beta_2 = \beta_{2\diamond}$ in the model $y = X_1\beta_1 + X\beta_2 + \varepsilon$ while ignoring β_2, we use

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

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

By specialising the expression under (15), a statistic may be derived for testing the hypothesis that $\beta_i = \beta_{i\diamond}$, concerning a single element:

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

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

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

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

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

The Decomposition of a Chi-Square Variate: Cochrane's Theorem

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

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

where the second equation comes from subtracting the unknown mean vector $X\beta$ from both sides of the first. In terms of the projector $P = X(X'X)^{-1}X'$, there is $X\hat{\beta} = Py$ and $e = y - X\hat{\beta} = (I - P)y$. Also, $\varepsilon = y - X\beta$. Therefore, the two equations can be written as

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

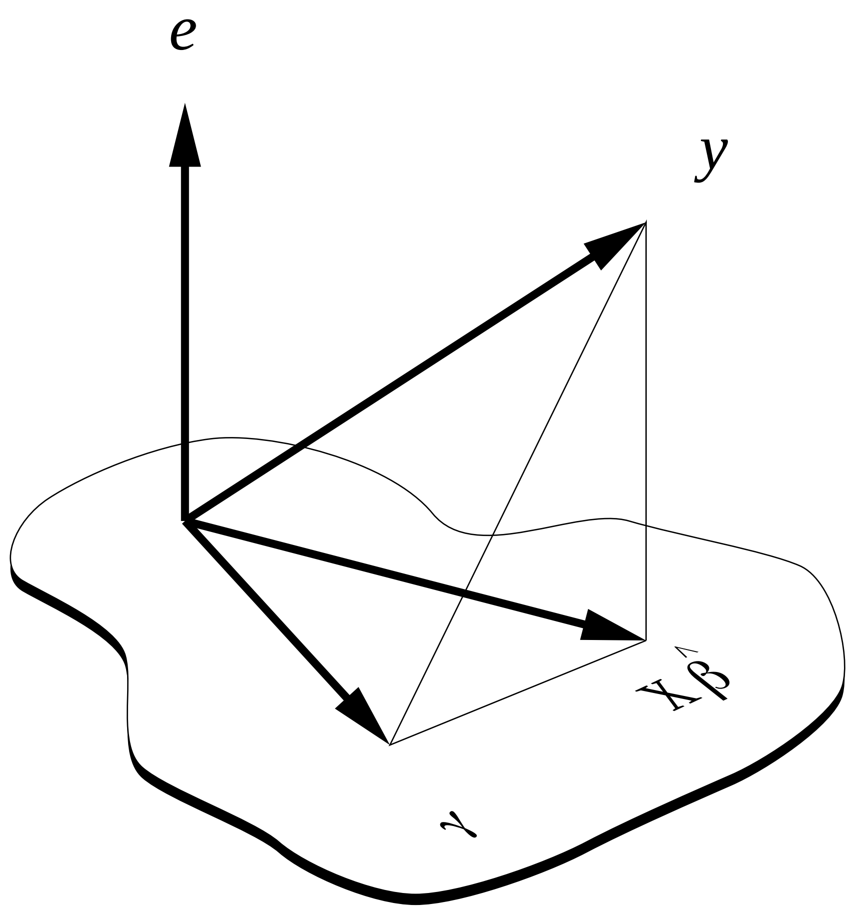

Figure 2. The vector $Py = X\hat{\beta}$ is formed by the orthogonal projection of the vector y onto the subspace spanned by the columns of the matrix X. Here, $\gamma = X\beta_\diamond$ is considered to be the true value of the mean vector.

From the fact that $P = P' = P^2$ and that $P'(I - P) = 0$, it follows 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 usual test of an hypothesis regarding the elements of the vector β is based on these relationships, which are depicted in Figure 2.

To test the hypothesis that $\beta_\diamond$ is the true value, the hypothesised mean $X\beta_\diamond$ is compared with the estimated mean vector $X\hat{\beta}$. The distance that separates the vectors is

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

This compared with the estimated variance of the disturbance term:

$$
(22)\qquad \hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T-k} = \frac{\varepsilon'(I-P)\varepsilon}{T-k},
$$

of which the numerator is the squared length of $e = (I - P)y = (I - P)\varepsilon$.

The arguments of the previous section, demonstrate that

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

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 (14) which is

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

Cochrane's Theorem

(25) Let $\varepsilon \sim N(0, \sigma^2 I_T)$ be a random vector of T independently and identically distributed elements. Also, let $P = X(X'X)^{-1}X'$ where X is of order $T \times k$ with 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.

Proof. To find an alternative expression for $P = X(X'X)^{-1}X'$, consider a matrix T such that $T(X'X)T' = I$ and $T'T = (X'X)^{-1}$. Then, $P = XT'TX' = C_1C_1'$, 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

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

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 $C'\varepsilon \sim N(0, \sigma^2 I_T)$. 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 $C_1 C_1' + C_2 C_2' = I_T$, the sum of these two variates is

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

and thus the theorem is proved.

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

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

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 the restrictions $R\beta = r$ on the regression coefficients of the model $N(y; X\beta, \sigma^2 I)$, where R is a $j \times k$ matrix rank j. Given that $\hat{\beta} \sim N\{\beta, \sigma^2(X'X)^{-1}\}$, it follows that

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

and, from this, it can be inferred immediately that

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

Since, it is statistically independent of $\hat{\beta}$,

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

must be statistically independent of the chi-square variate of (33).

Therefore, it can be deduced that

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

This F statistic can be used in testing the validity of the hypothesised restrictions $R\beta = r$.

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 $r = \beta_1^\diamond$.

The partitioned form of $(X'X)^{-1}$ is

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

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

With $R = [I, 0]$, we find that

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

Therefore, for testing the hypothesis that $\beta_1 = \beta_1^\diamond$, we use

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

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):

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

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

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

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

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

This equation can be rewritten as

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

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

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

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

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

On the other hand, there is

$$(48) \qquad RSS - USS = y'(I - P_2)y - y'(I - P)y = y'(P - P_2)y,$$

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:

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

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.

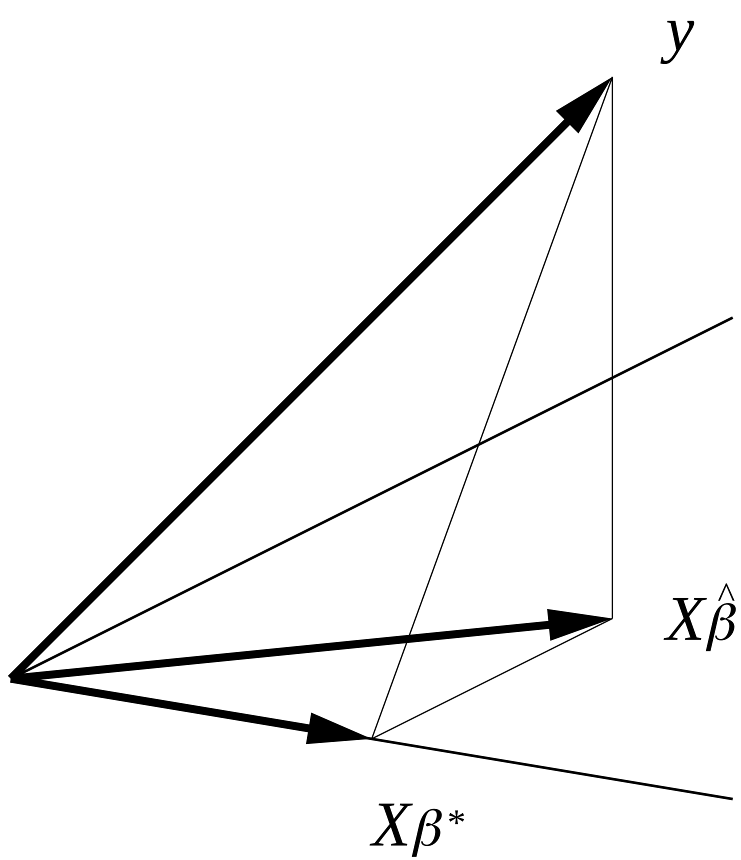

Figure 3. The test of the hypothesis entailed by the restricted model is based on a measure of the proximity of the restricted estimate $X\beta^*$, and the unrestricted estimate $X\hat{\beta}$. The USS is the squared distance $\|y - X\hat{\beta}\|^2$. The RSS is the squared distance $\|y - X\beta^*\|^2$.