Chapter 6

Bivariate Least Squares

6.1 The Bivariate Linear Model

Consider the model

$$y_i = \alpha + \beta x_i + u_i, \; i = 1, 2, \ldots, n, \tag{6.1}$$

where y_i is the dependent variable, x_i is the explanatory variable, and u_i is the unobservable disturbance. In matrix form, we have

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \tag{6.2}$$

or, more compactly, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. This is the *linear regression model*. It is linear in the variables (given α and β), linear in the parameters (given x_i), and linear in the disturbances.

Example 6.1 Neither of the following two equations is linear.

$$y_i = (\, \alpha + \beta x_i \,)\, u_i \tag{6.3}$$

$$y_i = \alpha x_i^{\beta} + u_i \tag{6.4}$$

□

6.1.1 Assumptions

For the disturbances, we assume that

(i) $\mathrm{E}(u_i) = 0$, for all i

(ii) $\mathrm{E}(u_i^2) = \sigma^2$, for all i

(iii) $\mathrm{E}(u_i u_j) = 0$, for all $i \neq j$

For the independent variable, We suppose

(iv) x_i nonstochastic for all i

(v) x_i nonconstant

For purposes of inference in finite samples, we sometime assume

(vi) $u_i \overset{iid}{\backsim} \mathrm{N}(0, \sigma^2)$, for all i.

6.1.2 Line Fitting

Consider a scatter of plots as shown in Figure 6.1.

x	y
2	12
3	7
4	8
5	5
6	3

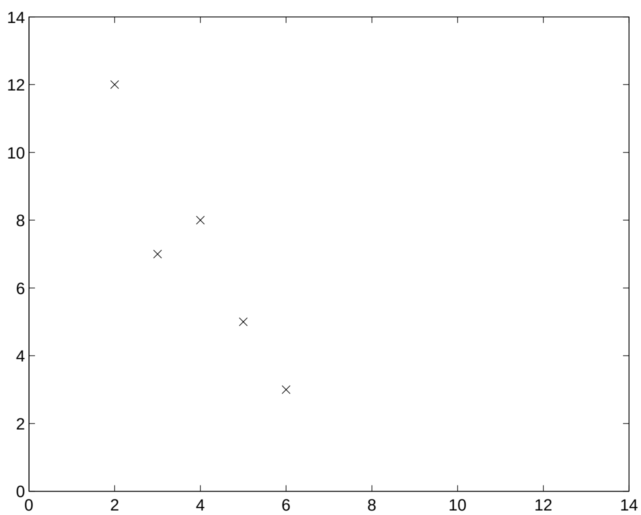

Figure 6.1: Scatter plot

We assume that x and y are linearly related. That is,

$$y_i = \alpha + \beta x_i. \tag{6.5}$$

Of course, we see that no single line "matches" all the observations.

Since no single line is entirely consistent with the data, we might choose α and β that best "fits" the data, in some sense. Define

$$u_i = y_i - (\alpha + \beta x_i), \text{ for } i = 1, 2, \ldots, n \tag{6.6}$$

as the discrepancy between the line choosen and the observations. Our objective then is to choose α and β to minimize the discrepancy.

A possible criterion for minimum discrepancy is

$$\min_{\alpha,\beta} \left| \sum_i u_i \right|, \tag{6.7}$$

but this will be zero for any line passing through $(\overline{x}, \overline{y})$. Another possibility is

$$\min_{\alpha,\beta} \sum_i |u_i|, \tag{6.8}$$

which is called the minimum absolute distance (MAD) or L_1 estimator. The MAD estimator has problems since the mathematics (and statistical distributions) are intractible. A closely related choice is

$$\min_{\alpha,\beta} \sum_i u_i^2. \tag{6.9}$$

This yields the least squares or L_2 estimator.

6.2 Least Squares Regression

6.2.1 The First-Order Conditions

Let

$$\phi = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (\, y_i - \alpha - \beta x_i \,)^2 \tag{6.10}$$

Then the minimum values, $\widehat{\alpha}$ and $\widehat{\beta}$ say, must satisfy the following first-order conditions:

$$0 = \frac{\partial \phi}{\partial \alpha} = \sum_{i=1}^{n} -2\,(\, y_i - \alpha - \beta x_i \,) \tag{6.11}$$

$$0 = \frac{\partial \phi}{\partial \beta} = \sum_{i=1}^{n} -2\,(\, y_i - \alpha - \beta x_i \,)\, x_i \tag{6.12}$$

Now, these first-order conditions may be written as

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (\, \alpha + \beta x_i \,) \tag{6.13}$$

$$\sum_{i=1}^{n} y_i x_i = \sum_{i=1}^{n} (\, \alpha x_i + \beta x_i^2 \,) \tag{6.14}$$

Thus, (??) implies

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}, \tag{6.15}$$

where $\overline{y} = \sum_{i=1}^{n} y_i/n$ and $\overline{x} = \sum_{i=1}^{n} x_i/n$.

Substituting (??) into (??) yields

$$\begin{aligned} \sum_{i=1}^{n} y_i x_i &= \sum_{i=1}^{n} [\, (\, \overline{y} - \widehat{\beta}\overline{x} \,)\, x_i + \beta x_i^2 \,] \\ &= \overline{y} \sum_{i=1}^{n} x_i + \widehat{\beta} \sum_{i=1}^{n} x_i \,(\, x_i - \overline{x} \,) \end{aligned} \tag{6.16}$$

and

$$\begin{aligned} \sum_{i=1}^{n} y_i x_i - n\overline{y}\overline{x} &= \widehat{\beta} \left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \right) \\ \sum_{i=1}^{n} (\, y_i - \overline{y} \,)(\, x_i - \overline{x} \,) &= \widehat{\beta} \sum_{i=1}^{n} (\, x_i - \overline{x} \,)^2 \end{aligned} \tag{6.17}$$

After solving for $\widehat{\beta}$, we have

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} (\, y_i - \overline{y} \,)(\, x_i - \overline{x} \,)}{\sum_{i=1}^{n} (\, x_i - \overline{x} \,)^2} \tag{6.18}$$

6.2.2 The Second-Order Conditions

Note that the least squares estimators $\widehat{\beta}$, and in turn, $\widehat{\alpha}$, are unique. Taking the second derivatives of the objective function yields

$$\frac{\partial^2 \phi}{\partial \alpha^2} = 2n, \tag{6.19}$$

$$\frac{\partial^2 \phi}{\partial \beta^2} = 2 \sum_{i=1}^{n} x_i^2, \tag{6.20}$$

and

$$\frac{\partial^2 \phi}{\partial \alpha \partial \beta} = 2 \sum_{i=1}^{n} x_i. \tag{6.21}$$

Thus, the Hessian matrix is

$$\mathbf{H} = \begin{pmatrix} 2n & 2\sum_{i=1}^{n} x_i \\ 2\sum_{i=1}^{n} x_i & 2\sum_{i=1}^{n} x_i^2 \end{pmatrix}, \tag{6.22}$$

which is a positive definite matrix, so $\widehat{\alpha}$ and $\widehat{\beta}$ are, in fact, minimums.

6.2.3 Matrix Interpretation

Now, the first-order conditions require (**??**) and (**??**), or

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (\,\widehat{\alpha} + \widehat{\beta} x_i\,) \tag{6.23}$$

$$\sum_{i=1}^{n} y_i x_i = \sum_{i=1}^{n} (\,\widehat{\alpha} x_i + \widehat{\beta} x_i^2\,) \tag{6.24}$$

These are the normal equations, and are linear in $\widehat{\alpha}$ and $\widehat{\beta}$. In matrix form, we have

$$\begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix}, \tag{6.25}$$

which has the solution

$$\begin{aligned}
\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} &= \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{pmatrix} \\
&= \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{pmatrix} \\
&= \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i \\ \\ -\sum_{i=1}^n x_i \sum_{i=1}^n y_i + n\sum_{i=1}^n x_i y_i \end{pmatrix}.
\end{aligned}$$

Now, $\widehat{\beta}$, according to this formula, is

$$\begin{aligned}
\widehat{\beta} &= \frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\
&= \frac{\sum_{i=1}^n x_i y_i - \overline{x}\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \overline{x}\sum_{i=1}^n x_i} \\
&= \frac{\sum_{i=1}^n x_i y_i - n\overline{xy}}{\sum_{i=1}^n x_i^2 - n\overline{x}^2} \\
&= \frac{\sum_{i=1}^n (\,y_i - \overline{y}\,)(\,x_i - \overline{x}\,)}{\sum_{i=1}^n (\,x_i - \overline{x}\,)^2},
\end{aligned} \tag{6.26}$$

while

$$\begin{aligned}
\widehat{\alpha} &= \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \qquad (6.27)\\
&= \frac{\overline{y} \sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i}\\
&= \frac{\overline{y} \sum_{i=1}^n x_i^2 - \overline{yx} \sum_{i=1}^n x_i + \overline{yx} \sum_{i=1}^n x_i - \overline{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i}\\
&= \frac{\overline{y}(\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i) + \overline{x}^2 \sum_{i=1}^n y_i - \overline{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i}\\
&= \overline{y} + \frac{\overline{x}^2 \sum_{i=1}^n y_i - \overline{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i}\\
&= \overline{y} + \frac{(\overline{x} \sum_{i=1}^n y_i - \sum_{i=1}^n x_i y_i)}{\sum_{i=1}^n x_i^2 - \overline{x} \sum_{i=1}^n x_i}\overline{x}\\
&= \overline{y} - \widehat{\beta}\overline{x}.
\end{aligned}$$

6.3 Basic Statistical Properties

6.3.1 Method Of Moments Interpretation

Suppose, for the moment, that x_i is a random variable that is uncorrelated with u_i. Let $\mu_x = \mathrm{E}(x_i)$. Then

$$\mu_y = \mathrm{E}(y_i) = \mathrm{E}(\alpha + \beta x_i + u_i) = \alpha + \beta\mu_x. \qquad (6.28)$$

Thus,

$$E(y_i - \mu_y) = \alpha + \beta x_i + u_i - (\alpha + \beta\mu_x) = \beta(x_i - \mu_x) + u_i, \qquad (6.29)$$

and

$$\mathrm{E}(y_i - \mu_y)(x_i - \mu_x) = \beta\, \mathrm{E}(x_i - \mu_x)^2 + \mathrm{E}(x_i - \mu_x)u_i = \beta\, \mathrm{E}(x_i - \mu_x)^2, \qquad (6.30)$$

since x_i and u_i are uncorrelated. Solving for β, we have

$$\beta = \frac{\mathrm{E}(x_i - \mu_x)(y_i - \mu_y)}{\mathrm{E}(x_i - \mu_x)^2}, \qquad (6.31)$$

while

$$\alpha = \mu_y - \beta\mu_x. \qquad (6.32)$$

Thus the least-squares estimators are method of moments estimators with sample moments replacing the population moments.

6.3.2 Mean of Estimates

Next, note that

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(\alpha + \beta x_i + u_i)}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\
&= \alpha\frac{\sum_{i=1}^{n}(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} + \beta\frac{\sum_{i=1}^{n}(x_i - \overline{x})x_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} + \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\
&= \beta + \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \qquad (6.33)
\end{aligned}$$

So, $\mathrm{E}(\widehat{\beta}) = \beta$, since $\mathrm{E}(u_i) = 0$ for all i. Thus, $\widehat{\beta}$ is an unbiased estimator of β. Further,

$$\begin{aligned}
\widehat{\alpha} &= \overline{y} - \beta\overline{x} \qquad (6.34) \\
&= \sum_{i=1}^{n}\frac{y_i}{n} - \overline{x}\frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \\
&= \sum_{i=1}^{n}\left(\frac{1}{n} - \overline{x}\frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)y_i \\
&= \alpha\sum_{i=1}^{n}\left(\frac{1}{n} - \overline{x}\frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right) \\
&\quad +\beta\sum_{i=1}^{n}\left(\frac{x_i}{n} - \overline{x}\frac{(x_i - \overline{x})x_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right) + \sum_{i=1}^{n}\left(\frac{1}{n} - \overline{x}\frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)u_i \\
&= \alpha\sum_{i=1}^{n}\left(\frac{1}{n} - \overline{x}\frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)u_i
\end{aligned}$$

So we have $\mathrm{E}(\widehat{\alpha}) = \alpha$, and $\widehat{\alpha}$ is an unbiased estimator of α.

6.3.3 Variance of Estimates

Now,

$$\widehat{\beta} - \beta = \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sum_{i=1}^{n} w_i u_i, \qquad (6.35)$$

where

$$w_i = \frac{(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \qquad (6.36)$$

So,

$$\begin{aligned}
\mathrm{Var}(\,\widehat{\beta}\,) &= \mathrm{E}(\,\widehat{\beta} - \beta\,)^2 \\
&= \mathrm{E}(\sum_{i=1}^{n} w_i u_i\,)^2 = \mathrm{E}(\,w_1 u_1 + w_2 u_2 + \cdots + w_n u_n\,)^2 \\
&= \mathrm{E}[(\,w_1^2 u_1^2 + w_2^2 u_2^2 + \cdots + w_n^2 u_n^2) + (w_1 u_1 w_2 u_2 + \cdots w_{n-1} u_{n-1} w_n u_n\,)] \\
&= w_1^2 \sigma^2 + w_2^2 \sigma^2 + \cdots + w_n^2 \sigma^2 \\
&= \sigma^2 \sum_{i=1}^{n} w_i^2 = \sigma^2 \sum_{i=1}^{n} \left(\frac{(\,x_i - \overline{x}\,)}{\sum_{j=1}^{n}(\,x_j - \overline{x}\,)^2} \right)^2 \\
&= \frac{\sigma^2}{\sum_{i=1}^{n}(\,x_i - \overline{x}\,)^2}. \qquad (6.37)
\end{aligned}$$

Next, we note that

$$\widehat{\alpha} - \alpha = \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{x} \frac{x_i - \overline{x}}{\sum_{i=1}^{n}(\,x_i - \overline{x}\,)^2} \right) u_i = \sum_{i=1}^{n} v_i u_i, \qquad (6.38)$$

where

$$v_i = \left(\frac{1}{n} - \overline{x} \frac{x_i - \overline{x}}{\sum_{i=1}^{n}(\,x_i - \overline{x}\,)^2} \right). \qquad (6.39)$$

So, in a fashion similar to the one above, we find that

$$\mathrm{Var}(\,\widehat{\alpha} - \alpha\,) = \sigma^2 \frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n}(\,x_i - \overline{x}\,)^2}, \qquad (6.40)$$

and

$$\mathrm{Cov}(\,\widehat{\alpha}, \widehat{\beta}\,) = \mathrm{E}(\,\widehat{\alpha} - \alpha\,)(\,\widehat{\beta} - \beta\,) = \sigma^2 \frac{\sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n}(\,x_i - \overline{x}\,)^2}. \qquad (6.41)$$

6.3.4 Estimation of σ^2

Next, we would like to get an estimate of σ^2. Let

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2, \qquad (6.42)$$

where

$$\begin{aligned} e_i &= y_i - \alpha - \beta x_i \\ &= (y_i - \overline{y}) - (\alpha - \alpha) - \widehat{\beta}(x_i - \overline{x}) \\ &= (y_i - \overline{y}) - \widehat{\beta}(x_i - \overline{x}) \\ &= \beta(x_i - \overline{x}) + (u_i - \overline{u}) - \widehat{\beta}(x_i - \overline{x}) \\ &= -(\widehat{\beta} - \beta)(x_i - \overline{x}) + (u_i - \overline{u}) \qquad (6.43) \\ & . \qquad (6.44) \end{aligned}$$

So,

$$\begin{aligned} \sum_{i=1}^n e_i^2 &= \sum_{i=1}^n [-(\widehat{\beta} - \beta)(x_i - \overline{x}) + (u_i - \overline{u})]^2 \qquad (6.45) \\ &= (\widehat{\beta} - \beta)^2 \sum_{i=1}^n (x_i - \overline{x})^2 \\ & \quad -2(\widehat{\beta} - \beta)\sum_{i=1}^n (x_i - \overline{x})(u_i - \overline{u}) + \sum_{i=1}^n (u_i - \overline{u})^2. \end{aligned}$$

Now, we have

$$\mathrm{E}[(\widehat{\beta} - \beta)^2 \sum_{i=1}^n (x_i - \overline{x})^2] = \sigma^2, \qquad (6.46)$$

$$\begin{aligned} \mathrm{E}(\sum_{i=1}^n (u_i - \overline{u})^2) &= \mathrm{E}(\sum_{i=1}^n (u_i^2 - \overline{u}^2)) \\ &= \mathrm{E}(\sum_{i=1}^n u_i^2) - \frac{1}{n}\mathrm{E}(\sum_{i=1}^n u_i)^2 = (n-1)\sigma^2, \qquad (6.47) \end{aligned}$$

and

$$\begin{aligned} & \mathrm{E}((\widehat{\beta} - \beta)\sum_{i=1}^n (x_i - \overline{x})(u_i - \overline{u})) \qquad (6.48) \\ &= \mathrm{E}\left(\sum_{i=1}^n w_i u_i\right)\left(\sum_{i=1}^n (x_i - \overline{x})u_i - \sum_{i=1}^n (x_i - \overline{x})\overline{u}\right) \\ &= \mathrm{E}\left(\sum_{i=1}^n w_i u_i\right)\sum_{i=1}^n (x_i - \overline{x})u_i = \sum_{i=1}^n w_i(x_i - \overline{x})\sigma^2 = \sigma^2. \end{aligned}$$

Therefore,

$$\mathrm{E}(\sum_{i=1}^n e_i^2) = \sigma^2 - 2\sigma^2 + (n-1)\sigma^2 = (n-2)\sigma^2 \qquad (6.49)$$

and

$$\mathrm{E}(\, s^2\,) = \sigma^2. \tag{6.50}$$

6.4 Statistical Properties Under Normality

6.4.1 Distribution of $\widehat{\beta}$

Suppose that $u_i \overset{iid}{\backsim} \mathrm{N}(\, 0, \sigma^2\,)$. Then,

$$\widehat{\beta} = \beta + \sum_{i=1}^{n} w_i u_i, \tag{6.51}$$

where $w_i = \frac{x_i - \overline{x}}{\sum_{i=1}^{n}(\, x_i - \overline{x}\,)^2}$. Then $\widehat{\beta}$ is also a normal random variable. Specifically,

$$\widehat{\beta} \sim \mathrm{N}(\, \beta, \sigma^2 q\,), \tag{6.52}$$

where $q = \frac{1}{\sum_{i=1}^{n}(\, x_i - \overline{x}\,)^2}$. Thus, for a given β_0,

$$\widehat{\beta} - \beta_0 \sim \mathrm{N}(\, \beta - beta_0, \sigma^2 q\,), \tag{6.53}$$

and

$$z_\beta = \frac{\widehat{\beta} - \beta_0}{\sqrt{\sigma^2 q}} \sim \mathrm{N}\left(\frac{\widehat{\beta} - \beta_0}{\sqrt{\sigma^2 q}}, 1 \right). \tag{6.54}$$

Now, suppose that $\mathrm{H}_0 : \beta = \beta_0$. Then,

$$\frac{\widehat{\beta} - \beta_0}{\sqrt{\sigma^2 q}} \sim \mathrm{N}(\, 0, 1\,), \tag{6.55}$$

while for $\mathrm{H}_0 : \beta = \beta_1 > \beta_0$ we have

$$\frac{\widehat{\beta} - \beta_0}{\sqrt{\sigma^2 q}} \sim \mathrm{N}\left(\frac{\beta_1 - \beta_0}{\sqrt{\sigma^2 q}}, 1 \right), \tag{6.56}$$

and we would expect the statistic to be centered to the right of zero.

6.4.2 Distribution of $\widehat{\alpha}$

Again, suppose that $u_i \overset{iid}{\backsim} \mathrm{N}(\, 0, \sigma^2\,)$. Then,

$$\widehat{\alpha} = \alpha + \sum_{i=1}^{n} v_i u_i, \tag{6.57}$$

where $v_i = \frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$. Then $\widehat{\alpha}$ is a normal random variable. Specifically,

$$\widehat{\alpha} \sim \mathrm{N}(\alpha, \sigma^2 p), \tag{6.58}$$

where $p = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \overline{x})^2}$. Then, for a given α_0,

$$\widehat{\alpha} - \alpha_0 \sim \mathrm{N}(\alpha - beta_0, \sigma^2 p), \tag{6.59}$$

and

$$z_\alpha = \frac{\widehat{\alpha} - \alpha_0}{\sqrt{\sigma^2 p}} \sim \mathrm{N}\left(\frac{\widehat{\alpha} - \alpha_0}{\sqrt{\sigma^2 p}}, 1\right). \tag{6.60}$$

Now, suppose that $\mathrm{H}_0 : \alpha = \alpha_0$. Then,

$$\frac{\widehat{\alpha} - \alpha_0}{\sqrt{\sigma^2 p}} \sim \mathrm{N}(0, 1), \tag{6.61}$$

while for $\mathrm{H}_0 : \alpha = \alpha_1 \neq \alpha_0$ we have

$$\frac{\widehat{\alpha} - \alpha_0}{\sqrt{\sigma^2 p}} \sim \mathrm{N}\left(\frac{\alpha_1 - \alpha_0}{\sqrt{\sigma^2 p}}, 1\right), \tag{6.62}$$

and we would expect the statistic not to be centered around zero.

It is important to note that

$$p = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \overline{x})^2},$$

and so $\sqrt{\sigma^2 p}$ grows small as n grows large. Thus, the noncentrality of the distribution of the statistic z_α will also grow under the alternative. A similar statemetn can be made concerning q and the statistic z_β.

6.4.3 t-Distribution

In most cases, we do not know the value of σ^2. A possible alternative is to use s^2, whereupon we obtain

$$\frac{\widehat{\alpha} - \alpha_0}{\sqrt{s^2 p}} \sim \mathrm{t}_{n-2}, \tag{6.63}$$

under $\mathrm{H}_0 : \alpha = \alpha_0$, and

$$\frac{\widehat{\beta} - \beta_0}{\sqrt{s^2 q}} \sim \mathrm{t}_{n-2}, \tag{6.64}$$

under $H_0 : \beta = \beta_0$.

The t-distibution is quite similar to the standard normal except being slightly fatter. This reflects the added uncertainty introduced by using s^2 rather than σ^2. However, as n increases and the precision of s^2 becomes better, the t-distibution grows closer and closer to the $N(0, 1)$. We loose two degrees of freedom (and so t_{n-2}) because of the fact that we estimated two coefficients, namely α and β.

Just as in the case of z_α and z_β, we would expect the t-distribution to be off-center if the null hypothesis were not true.

6.4.4 Maximum Likelihood

Suppose that $u_i \overset{iid}{\backsim} N(0, \sigma^2)$. Then,

$$y_i \overset{iid}{\backsim} N(\alpha + \beta x_i, \sigma^2). \tag{6.65}$$

Then, the pdf of y_i is given by

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} [y_i - (\alpha + \beta x_i)]^2 \right\}. \tag{6.66}$$

Since the observations are independent, we can write the joint likelihood function as

$$\begin{aligned} f(y_1, y_2, \ldots, y_n) &= f(y_1) f(y_2) \cdots f(y_n) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - (\alpha + \beta x_i)]^2 \right\} \\ &= L(\alpha, \beta, \sigma^2 | \mathbf{y}, \mathbf{x}). \end{aligned} \tag{6.67}$$

Now, let $\mathcal{L} = \log L(\alpha, \beta, \sigma^2 | \mathbf{y}, \mathbf{x})$. We seek to maximize

$$\mathcal{L} = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - (\alpha + \beta x_i)]^2. \tag{6.68}$$

Note that for (**??**) to be a maximum with respect to α and β, we most minimize $\sum_{i=1}^{n} [y_i - (\alpha + \beta x_i)]^2$.

The first-order conditions for (**??**) are

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \frac{1}{\sigma^2} \sum_{i=1}^{n} [y_i - (\widehat{\alpha} + \widehat{\beta} x_i)] = 0, \tag{6.69}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = \frac{1}{\sigma^2} \sum_{i=1}^{n} [y_i - (\widehat{\alpha} + \widehat{\beta} x_i)\, x_i] = 0, \tag{6.70}$$

x_i	y_i	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
2	12	-2	5	4	-24
3	7	-1	0	1	-7
4	8	0	1	0	0
5	5	1	-2	1	5
6	3	2	-4	4	6
20	35	0	0	10	-20

Table 6.1: Summary table.

and

$$\frac{\partial \mathcal{L}}{\partial \beta} = -\frac{n}{2\widehat{\sigma^2}} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}[\, y_i - (\, \alpha + \beta x_i \,)]^2. \tag{6.71}$$

Note that the first two conditions imply that

$$\overline{\alpha} = \overline{y} - \widehat{\beta}\overline{x}, \tag{6.72}$$

and

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(\, x_i - \overline{x}\,)(\, y_i - \overline{y}\,)}{\sum_{i=1}^{n}(\, x_i - \overline{x}\,)^2}, \tag{6.73}$$

since these are the same as the normal equations (except for σ^2). The third condition yields

$$\begin{aligned} \widehat{\sigma^2} &= \frac{\sum_{i=1}^{n}[\, y_i - (\, \alpha + \beta x_i \,)]^2}{n} \\ &= \frac{\sum_{i=1}^{n} e_i^2}{n} = \frac{n-2}{n}s^2. \end{aligned} \tag{6.74}$$

6.5 An Example

Consider the scatter graph given in Figure 6.1. From this, we construct Table 6.1.

Thus, we have

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(\, x_i - \overline{x}\,)(\, y_i - \overline{y}\,)}{\sum_{i=1}^{n}(\, x_i - \overline{x}\,)^2} = \frac{-20}{10} = -2,$$

and

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x} = 7 - (\, -2\,)4 = 15.$$

$\widehat{\beta}x_i$	$\widehat{\alpha}+\widehat{\beta}x_i$	e_i	e_i^2
-4	11	1	1
-6	9	-2	4
-8	7	1	1
-10	5	0	0
-12	3	0	0

Table 6.2: Residual calculations.

Now, we calculate the residuals in Table 6.2, and find

$$\sum_{i=1}^{n} e_i^2 = 6$$

and

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{6}{3} = 2.$$

An estimate of the variance of $\widehat{\beta}$, namely $\sigma^2 q$, is provided by

$$s^2 q = s^2 \frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 2\frac{1}{10} = 0.2.$$

Suppose we wish to test $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$. Then

$$\frac{\widehat{\beta} - 0}{\sqrt{s^2 q}} \sim \mathrm{t}_3$$

under the null hypothesis, but

$$\frac{\widehat{\beta} - 0}{\sqrt{s^2 q}} = \frac{-2}{\sqrt{0.2}} = \frac{-2}{0.45} = -4.2$$

is clearly in the left-hand 2.5% tail of the t_3-distribution. Thus, we would reject the null hypothesis at the 95% significance level.

Chapter 7

Linear Least Squares

7.1 Multiple Regression Model

The general k-variable linear model can be written as

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i \qquad i = 1, 2, \ldots, n. \tag{7.1}$$

Using matrix techniques, we can equivalently write this model as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \tag{7.2}$$

or, more compactly, as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \tag{7.3}$$

7.1.1 Assumptions

In the general k-variable linear model, we make the following assumptions about the disturbances:

(i) $\mathrm{E}(\, u_i \,) = 0 \qquad i = 1, 2, \ldots, n$

(ii) $\mathrm{E}(\, u_i^2 \,) = 0 \qquad i = 1, 2, \ldots, n$

(iii) $\mathrm{E}(\, u_i u_j \,) = 0 \qquad i \neq j$

These assumptions can be written in matrix notaion as

$$\mathrm{E}(\, \mathbf{u} \,) = 0 \tag{7.4}$$

and

$$\text{Cov}(\mathbf{u}) = \text{E}(\mathbf{u}\mathbf{u}') = \sigma^2 \text{I}_n, \tag{7.5}$$

where I_n is an $n \times n$ identity matrix.

The nonstochastic assumptions are

(iv) $\mathbf{X}$ is nonstochastic.

(v) $\mathbf{X}$ has full column rank (the columns are linearly independent).

Sometimes, we will also assume that the u_i's are normally distrubuted

(vi) $\mathbf{u} \sim \text{E}(0, \sigma^2 \text{I}_n)$.

7.1.2 Plane Fitting

Suppose $k = 3$ and $x_{i1} = 1$. Then,

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i \qquad i = 1, 2, \ldots, n. \tag{7.6}$$

Now,

$$\widehat{y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \widehat{\beta}_3 x_{i3} \qquad i = 1, 2, \ldots, n. \tag{7.7}$$

define planes in the three-dimentional space of y, x_2, and x_3. We seek to choose $\widehat{\beta}_1, \widehat{\beta}_2$ and $\widehat{\beta}_3$ so that the points on the plane corresponding to x_{i2} and x_{i3}, namely $\widehat{y}_i$, will be close to y_i. That is, we will "fit" a plane to the observations.

As we did in the two-dimensional case, we choose to measure closeness in the vertical distance. That is,

$$e_i = y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \widehat{\beta}_3 x_{i3}) \qquad i = 1, 2, \ldots, n, \tag{7.8}$$

and

$$\phi = \sum_{i=1}^{n} e_t^2 \tag{7.9}$$

7.1.3 Least Squares

In general, we want to

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} [y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \cdots + \widehat{\beta}_k x_{ik})]^2 \tag{7.10}$$

7.2 Least Squares Regression

7.2.1 The OLS Estimator

As was stated above, we seek to minimize

$$\phi = \sum_{i=1}^{n} [\, y_i - (\, \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}\,)]^2 \tag{7.11}$$

with respect to the coefficients $\beta_1, \beta_2, \ldots, \beta_k$. The first-order conditions are

$$\begin{aligned}
0 = \frac{\partial \phi}{\partial \beta_1} &= 2\sum_{i=1}^{n} [\, y_i - (\, \widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \cdots + \widehat{\beta}_k x_{ik}\,)] x_{i1}, \qquad (7.12)\\
0 = \frac{\partial \phi}{\partial \beta_2} &= 2\sum_{i=1}^{n} [\, y_i - (\, \widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \cdots + \widehat{\beta}_k x_{ik}\,)] x_{i2},\\
&\vdots\\
0 = \frac{\partial \phi}{\partial \beta_k} &= 2\sum_{i=1}^{n} [\, y_i - (\, \widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \cdots + \widehat{\beta}_k x_{ik}\,)] x_{ik}.
\end{aligned}$$

where $(\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k)$ are solutions. Rearranging, we have the normal equations:

$$\begin{aligned}
\widehat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 + \widehat{\beta}_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} x_{i1}x_{ik} &= \sum_{i=1}^{n} x_{i1}y_i\\
\widehat{\beta}_1 \sum_{i=1}^{n} x_{i2}x_{i1} + \widehat{\beta}_2 \sum_{i=1}^{n} x_{i2}^2 + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} x_{i2}x_{ik} &= \sum_{i=1}^{n} x_{i2}y_i\\
&\vdots\\
\widehat{\beta}_1 \sum_{i=1}^{n} x_{ik}x_{i1} + \widehat{\beta}_2 \sum_{i=1}^{n} x_{ik}x_{i2} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 &= \sum_{i=1}^{n} x_{ik}y_i
\end{aligned} \tag{7.13}$$

or

$$\begin{pmatrix}
\sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik}\\
\sum_{i=1}^n x_{i2}x_{i1} & \sum_{i=1}^n x_{i2}^2 & \cdots & \sum_{i=1}^n x_{i2}x_{ik}\\
\vdots & \vdots & \ddots & \vdots\\
\sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2
\end{pmatrix}
\begin{pmatrix} \widehat{\beta}_1\\ \widehat{\beta}_2\\ \vdots\\ \widehat{\beta}_k \end{pmatrix}
=
\begin{pmatrix} \sum_{i=1}^n x_{i1}y_i\\ \sum_{i=1}^n x_{i2}y_i\\ \vdots\\ \sum_{i=1}^n x_{ik}y_i \end{pmatrix}. \tag{7.14}$$

As in the bivariate model,

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix}, \tag{7.15}$$

so

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \sum_{i=1}^n x_{i2}x_{i1} & \sum_{i=1}^n x_{i2}^2 & \cdots & \sum_{i=1}^n x_{i2}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{pmatrix}, \tag{7.16}$$

and

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} \sum_{i=1}^n x_{i1}y_i \\ \sum_{i=1}^n x_{i2}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{pmatrix}. \tag{7.17}$$

Thus, we can write the normal equations in matrix notation:

$$\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}. \tag{7.18}$$

where $\boldsymbol{\beta}' = (\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k)$.Therefore, we have the unique solution

$$\widehat{\boldsymbol{\beta}} = (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{y}, \tag{7.19}$$

as long as $|\mathbf{X}'\mathbf{X}| \neq 0$, which is assured by Assumption (v).

7.2.2 Some Algebraic Results

Define the fitted value for each i as $\widehat{y}_i = x_{i1}\widehat{\beta}_1 + x_{i2}\widehat{\beta}_2 + \cdots + x_{ik}\widehat{\beta}_k = \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$ whereupon

$$\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}. \tag{7.20}$$

Next define the OLS residual for each i as $e_i = y_i - \widehat{y}_i$ so

$$\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}}. \tag{7.21}$$

Then,

$$\begin{aligned} \mathbf{X}'\mathbf{e} &= \mathbf{X}'(\,\mathbf{y} - \widehat{\mathbf{y}}\,) \\ &= \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{0}, \end{aligned} \tag{7.22}$$

and we say that the residuals are orthogonal to the regressors. Also, we find

$$\begin{aligned}
\widehat{\mathbf{y}}'\mathbf{y} &= (\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{e}) \\
&= \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{e} \\
&= \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} \\
&= \widehat{\mathbf{y}}'\widehat{\mathbf{y}}. \qquad (7.23)
\end{aligned}$$

Now, suppose that the first coefficient is the intercept. Then, the first column of $\mathbf{X}$ and hence the first row of $\mathbf{X}'$ are all ones. This means that

$$\begin{aligned}
0 = \mathbf{X}'\mathbf{e} &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n e_i \\ \sum_{i=1}^n x_{i2}e_i \\ \vdots \\ \sum_{i=1}^n x_{ik}e_i \end{pmatrix}. \qquad (7.24)
\end{aligned}$$

So, $\sum_{i=1}^n e_i = 0$, which means that

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \widehat{y}_i + e_i = \sum_{i=1}^n \widehat{y}_i. \qquad (7.25)$$

Finally, we note that

$$\begin{aligned}
\mathbf{e} &= \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \\
&= \mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= [\,\mathrm{I}_n - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,]\mathbf{y} \\
&= \mathbf{M}\mathbf{y} \\
&= \mathbf{M}(\,\mathbf{X}\boldsymbol{\beta} + \mathbf{u}\,) \\
&= [\,\mathrm{I}_n - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,](\,\mathbf{X}\boldsymbol{\beta}\,) + \mathbf{M}\mathbf{u} \\
&= \mathbf{M}\mathbf{u}. \qquad (7.26)
\end{aligned}$$

We see that the OLS residuals are a linear transformation of the underlying disturbances. The matrix M, which is sometimes called "the idempotent matrix" plays an important role in the sequel, has the property $M = M \cdot M$ or of idempotence.

7.2.3 The R^2 Statistic

Define the following:

$$\mathrm{SSE} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (\, y_i - \widehat{y}_i \,)^2, \tag{7.27}$$

$$\mathrm{SST} = \sum_{i=1}^{n} (\, y_i - \overline{y} \,)^2, \tag{7.28}$$

$$\mathrm{SSR} = \sum_{i=1}^{n} (\, \widehat{y}_i - \overline{y} \,)^2. \tag{7.29}$$

Note that SSE is the variation of actuals around the fitted plane and is called the unexplained sum-of-squares. SSR, or residual sum-of-squares, is variation of the fitted values around the sample mean and SST, or total sum-of-squares, is the variation of the actual around the sample mean.

The three sums-of-squares are closely related. Consider

$$\begin{aligned}
\mathrm{SST} - \mathrm{SSE} &= \sum_{i=1}^{n} [(\, y_i - \overline{y} \,)^2 - (\, y_i - \widehat{y}_i \,)^2] \\
&= \sum_{i=1}^{n} \overline{y}^2 - 2\overline{y} \sum_{i=1}^{n} y_i + 2 \sum_{i=1}^{n} y_i \widehat{y}_i - \sum_{i=1}^{n} \widehat{y}_i^2 \\
&= \sum_{i=1}^{n} \overline{y}^2 - 2\overline{y} \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \widehat{y}_i^2 \\
&= \sum_{i=1}^{n} \overline{y}^2 - 2\overline{y} \sum_{i=1}^{n} \widehat{y}_i + \sum_{i=1}^{n} \widehat{y}_i^2 \\
&= \sum_{i=1}^{n} (\, \widehat{y}_i - \overline{y} \,)^2 = \mathrm{SSR}.
\end{aligned} \tag{7.30}$$

Thus $\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}$.

We now define

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} = \frac{\mathrm{SST}}{\mathrm{SST}} - \frac{\mathrm{SSE}}{\mathrm{SST}} = \frac{\mathrm{SSR}}{\mathrm{SST}}. \tag{7.31}$$

as the percent of of total variation explained by the model.

This statistic can also be interpreted as a squared correlation coefficient. Consider the sample second moments,

$$\mathrm{Var}(\, \mathbf{y} \,) = \frac{1}{n} \sum_{i=1}^{n} (\, y_i - \overline{y} \,)^2 = \frac{1}{n} \mathrm{SST}, \tag{7.32}$$

$$\mathrm{Var}(\,\widehat{\mathbf{y}}\,) = \frac{1}{n}\sum_{i=1}^{n}(\,\widehat{y}_i - \overline{y}\,)^2 = \frac{1}{n}\mathrm{SSR}, \tag{7.33}$$

and

$$\begin{aligned}
\mathrm{Cov}(\,\mathbf{y},\widehat{\mathbf{y}}\,) &= \frac{1}{n}\sum_{i=1}^{n}(\,y_i - \overline{y}\,)(\,\widehat{y}_i - \overline{y}\,) \\
&= \frac{1}{n}\sum_{i=1}^{n} y_i\widehat{y}_i - \overline{y}\widehat{y}_i - \overline{y}y_i + \overline{y}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i^2 - 2\overline{y}\widehat{y}_i + \overline{y}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(\,\widehat{y}_i - \overline{y}\,)^2 \\
&= \frac{1}{n}\mathrm{SSR}.
\end{aligned} \tag{7.34}$$

Then the correlation between y_i and $\widehat{y}_i$ can be written as,

$$\begin{aligned}
\widehat{\rho}_{y,\widehat{y}} &= \frac{\mathrm{Cov}(\,\mathbf{y},\widehat{\mathbf{y}}\,)}{\sqrt{\mathrm{Var}(\,\widehat{\mathbf{y}}\,)\ \mathrm{Var}(\,\widehat{\mathbf{y}}\,)}} \\
&= \frac{\frac{1}{n}\mathrm{SSR}}{\sqrt{\frac{1}{n}\mathrm{SST}\frac{1}{n}\mathrm{SSR}}} \\
&= \frac{\sqrt{\mathrm{SSR}}}{\sqrt{\mathrm{SST}}},
\end{aligned}$$

so

$$\widehat{\rho}^2_{y,\widehat{y}} = \frac{\mathrm{SSR}}{\mathrm{SST}} = \mathrm{R}^2. \tag{7.35}$$

This statistic is variously called the coefficient of determination, the multiple correlation statistic and the "R"-squared statistic.

7.3 Basic Statistical Results

7.3.1 Mean and Covariance of $\widehat{\boldsymbol{\beta}}$

From the previous section, under assumption (v), we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{y} \\
&= (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'(\,\mathbf{X}\boldsymbol{\beta} + \mathbf{u}\,) \\
&= (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{u} \\
&= \boldsymbol{\beta} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{u}.
\end{aligned} \tag{7.36}$$

Since $\mathbf{X}$ is nonstochasticby assumption (iv), we have, also using assumption (i),

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' E(\mathbf{u}) \\ &= \boldsymbol{\beta}. \end{aligned} \tag{7.37}$$

Thus, the OLS estimator is unbiased. Also, using assumptions (ii) and (iii),

$$\begin{aligned} \mathrm{Cov}(\widehat{\boldsymbol{\beta}}) &= E(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2 \mathrm{I}_n \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \tag{7.38}$$

7.3.2 Best Linear Unbiased Estimator (BLUE)

The OLS estimator is linear in $\mathbf{y}$ and it is an unbiased estimator, as we saw above. Let

$$\widetilde{\boldsymbol{\beta}} = \widetilde{\mathbf{A}}\mathbf{y}, \tag{7.39}$$

where $\mathbf{A}$ is a $k \times n$ matrix that is nonstochastic, be any other unbiased estimator. That is, $E(\widetilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. Define

$$\mathbf{A} = \widetilde{\mathbf{A}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{7.40}$$

Then,

$$\begin{aligned} \widetilde{\boldsymbol{\beta}} &= [\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \\ &= [\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'][\mathbf{X}\boldsymbol{\beta} + \mathbf{u}] \\ &= \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta} + [\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{u}. \end{aligned} \tag{7.41}$$

Now, $\widetilde{\boldsymbol{\beta}}$ is an unbiased estimator, so

$$\begin{aligned} E(\widetilde{\boldsymbol{\beta}}) &= \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta} + [\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] E(\mathbf{u}) \\ &= \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta} = \boldsymbol{\beta}, \end{aligned} \tag{7.42}$$

which implies that for all $\boldsymbol{\beta}$, $\mathbf{A}\mathbf{X} = 0$. Thus,

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + [\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{u}, \tag{7.43}$$

and

$$\begin{aligned} \mathrm{Cov}(\widetilde{\boldsymbol{\beta}}) &= E(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})' \\ &= E\{[\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{u}\mathbf{u}'[\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']'\} \end{aligned}$$

$$\begin{aligned} &= [\,\mathbf{A} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\,]\,\mathrm{E}(\,\mathbf{uu}'\,)[\,\mathbf{A} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\,]' \\ &= [\,\mathbf{A} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\,]\sigma^2\mathrm{I}_n[\,\mathbf{A} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\,]' \\ &= \sigma^2[\,\mathbf{AA}' + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\,] \\ &= \sigma^2\mathbf{AA}' + \sigma^2(\,\mathbf{X}'\mathbf{X}\,)^{-1}. \end{aligned} \tag{7.44}$$

This shows that the covariance matrix of any other linear unbiased estimator exceeds the covariance matrix of the OLS estimator by a possitive semi-definite matrix $\sigma^2\mathbf{AA}'$. Hence, OLS is said to be best linear unbiased estimator (BLUE). Note that we have used all of the assumptions (i)-(v) to get to this point.

7.3.3 Consistancy

Typically, the elements of $\mathbf{X}'\mathbf{X}$ are unbounded (they go to infinity) as n gets very large. For example, the $1,1$ element is n and the j,j element is $\sum_{i=1}^{n} x_{ij}^2$. Therefore,

$$\lim_{n\to\infty} (\,\mathbf{X}'\mathbf{X}\,)^{-1} = 0, \tag{7.45}$$

and the variances of $\widehat{\boldsymbol{\beta}}$ converge to zero. This means that the distribution collapses about its expected value, namely $\boldsymbol{\beta}$. So,

$$\operatorname*{plim}_{n\to\infty} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}, \tag{7.46}$$

and OLS estimation is consistent. A more formal proof of this property will be given in the chapter on stochastic regressors.

7.3.4 Estimation Of σ^2

Recall that

$$\mathbf{e} = \mathbf{Mu}, \tag{7.47}$$

where $\mathbf{M} = \mathrm{I}_n - \mathbf{X}(\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'$. Then,

$$\mathbf{e}'\mathbf{e} = (\,\mathbf{Mu}\,)'\mathbf{Mu} = \mathbf{u}'\mathbf{MMu} = \mathbf{u}'\mathbf{Mu}, \tag{7.48}$$

since $\mathbf{M}$ is symmetric and idempotent. Also,

$$\mathbf{e}'\mathbf{e} = \mathrm{tr}\,\mathbf{e}'\mathbf{e} = \mathrm{tr}\,\mathbf{u}'\mathbf{Mu} = \mathrm{tr}\,\mathbf{Mu}'\mathbf{u}, \tag{7.49}$$

since $\mathbf{e}'\mathbf{e}$ is a scalar, and $\mathrm{tr}\,\mathbf{AB} = \mathrm{tr}\,\mathbf{BA}$, when both multiplications are defined. Thus,

$$\begin{aligned} \mathrm{E}(\,\mathbf{e}'\mathbf{e}\,) &= \mathrm{E}(\,\mathrm{tr}\,\mathbf{Mu}'\mathbf{u}\,) \\ &= \mathrm{tr}\,\mathbf{M}\,\mathrm{E}(\,\mathbf{u}'\mathbf{u}\,) \\ &= \mathrm{tr}\,\mathbf{M}\sigma^2\mathrm{I}_n \\ &= \sigma^2\,\mathrm{tr}\,\mathbf{M}. \end{aligned} \tag{7.50}$$

But,

$$\begin{aligned} \operatorname{tr}\mathbf{M} &= \operatorname{tr}(\mathrm{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= \operatorname{tr}\mathrm{I}_n - \operatorname{tr}(,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= n - \operatorname{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) \\ &= n - k. \end{aligned} \tag{7.51}$$

Now, define

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k}. \tag{7.52}$$

Then

$$\mathrm{E}(s^2) = \frac{\mathrm{E}(\mathbf{e}'\mathbf{e})}{n-k} = \frac{\sigma^2(n-k)}{n-k} = \sigma^2, \tag{7.53}$$

so s^2 is an unbiased estimator of σ^2. We can also establish that s^2 is a consistent estimator of σ^2. That is,

$$\operatorname*{plim}_{n\to\infty} s^2 = \sigma^2. \tag{7.54}$$

7.3.5 Prediction

Suppose that we wish to predict

$$y_p = \beta_1 x_{p1} + \beta_2 x_{p2} + \cdots + \beta_k x_{pk} + u_p = \mathbf{x}_p{}'\boldsymbol{\beta} + u_p. \tag{7.55}$$

Note that

$$\mathrm{E}(y_p|\mathbf{x}_p) = \mathbf{x}_p'\boldsymbol{\beta}. \tag{7.56}$$

A natural choice for a predictor is

$$\begin{aligned} \widehat{y}_p &= \mathbf{x}_p'\widehat{\boldsymbol{\beta}} \\ &= \mathbf{x}_p'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \mathbf{x}_p'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}'\boldsymbol{\beta} + \mathbf{u} \\ &= \mathbf{x}_p'\boldsymbol{\beta} + \mathbf{x}_p'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}. \end{aligned} \tag{7.57}$$

Now,

$$\mathrm{E}(y_p|\mathbf{x}_p) = \mathbf{x}_p^{'}\boldsymbol{\beta}, \tag{7.58}$$

and

$$\mathrm{E}[(y_p - \widehat{y}_p)|\mathbf{x}_p] = 0. \tag{7.59}$$

Hence, $\widehat{y}_p$ is an unbiased predictor of y_p.

We also have

$$\mathrm{Var}(\widehat{y}_p) = \mathrm{E}(\widehat{y}_p - \mathbf{x}_p^{'}\boldsymbol{\beta})^2 = \sigma^2\mathbf{x}_p^{'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_p, \tag{7.60}$$

while

$$\mathrm{MSPE}(\widehat{y}_p) = \mathrm{E}(y_p - \widehat{y}_p) = \sigma^2[1 + \mathbf{x}_p^{'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_p]. \tag{7.61}$$

It can be shown as above that $\widehat{y}_p$ is the best (minimum variance) linear unbiased predictor (BLUP) of y_p.

7.4 Statistical Properties Under Normality

7.4.1 Distribution Of $\widehat{\boldsymbol{\beta}}$

Suppose that we introduce assumption (vi), so the u_i's are normal:

$$\mathbf{u} \sim \mathrm{E}(\,0, \sigma^2 \mathrm{I}_n\,). \tag{7.62}$$

Recall that

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{u}, \tag{7.63}$$

so $\widehat{\boldsymbol{\beta}}$ is linear in $\mathbf{u}$ and $\mathbf{X}$ and hence $(\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'$ are nonstochastic. Then, $\widehat{\boldsymbol{\beta}}$ is also normally distributed:

$$\widehat{\boldsymbol{\beta}} \sim \mathrm{E}(\,0, \sigma^2(\,\mathbf{X}'\mathbf{X}\,)^{-1}\,). \tag{7.64}$$

Thus, we may test $\mathrm{H}_0 : \beta_i = \beta_i 0$ with

$$\frac{\widehat{\boldsymbol{\beta}} - \beta_i 0}{\sqrt{\sigma^2(\,\mathbf{X}'\mathbf{X}\,)^{-1}_{ii}}} \sim \mathrm{E}(\,0, 1\,). \tag{7.65}$$

More will be said on this statistic for use in inference in the next chapter.

7.4.2 Maximum Likelihood Estimation

Now,

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i = \mathbf{x}_i{}'\boldsymbol{\beta} + u_i, \tag{7.66}$$

is linear in u_i, so y_i is also normal given $\mathbf{x}_i$:

$$y_i \sim N(\,\mathbf{x}_i{}'\boldsymbol{\beta}, \sigma^2\,). \tag{7.67}$$

Further, the y_i's are independent. Thus, the density for y_i is given by

$$f(\,y_i\,) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}[\,y_i - \mathbf{x}_i{}'\boldsymbol{\beta}\,]^2 \right\}. \tag{7.68}$$

Since the y_i's are independent, the joint likelihood function is

$$\begin{aligned} f(\,y_1, y_2, \ldots, y_n\,) &= f(\,y_1\,)f(\,y_2\,)\cdots f(\,y_n\,) \\ &= \frac{1}{(\,2\pi\sigma^2\,)^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [\,y_i - \mathbf{x}_i{}'\boldsymbol{\beta}\,]^2 \right\} \\ &= \mathrm{L}(\,\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}\,). \end{aligned} \tag{7.69}$$

Let $\mathcal{L} = \log \mathrm{L}(\,\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}\,)$. We wish to maximize $\mathcal{L}$ with respect $\boldsymbol{\beta}$. However, this means that we minimize the sum of squares, so

$$\widehat{\boldsymbol{\beta}}_{MLE} = (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{y}, \tag{7.70}$$

which is the OLS estimator from above. It is easily shown that

$$\widehat{\sigma^2}_{MLE} = \frac{\mathbf{e}'\mathbf{e}}{n} = \frac{n-k}{n} s^2. \tag{7.71}$$

7.4.3 Efficiency of $\widehat{\boldsymbol{\beta}}$ and s^2

Since $\widehat{\boldsymbol{\beta}}$ is the MLE and unbiased, we find then it is the minimum variance unbiased estimator (BUE). And s^2 is not the MLE, so it is not BUE. On the other hand, $\widehat{\sigma^2}_{MLE}$ is biased, so it is not BUE either. The will both be equivalent in large samples and be asymptotically BUE.

Chapter 8

Confidence Intervals and Hypothesis Tests

8.1 Introduction

8.1.1 Model and Assumptions

The model is a k-variable linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{8.1}$$

where $\mathbf{y}$ and $\mathbf{u}$ are both $n \times 1$ vectors, $\mathbf{X}$ is a $n \times k$ matrix and $\boldsymbol{\beta}$ is a $k \times 1$ vector. We make the following assumptions about the disturbances:

(i) $\mathrm{E}(\mathbf{u}) = 0$

and

(ii),(iii) $\mathrm{Cov}(\mathbf{u}) = \mathrm{E}(\mathbf{u}\mathbf{u}') = \sigma^2 \mathrm{I}_n$,

where I_n is an $n \times n$ identity matrix. The nonstochastic assumptions are

(iv) $\mathbf{X}$ is nonstochastic.

(v) $\mathbf{X}$ has full column rank (the columns are linearly independent).

For inferences, we assume that $\mathbf{u}$ are normally distributed. That is,

(vi) $\mathbf{u} \sim (0, \sigma^2 \mathrm{I}_n)$.

8.1.2 Ordinary Least Squares Estimation

For some estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, define

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \tag{8.2}$$

and

$$\phi = \mathbf{e}'\mathbf{e} \tag{8.3}$$

Choosing $\widehat{\beta}$ to minimize ϕ yields the ordinary least squares (OLS) estimator

$$\widehat{\boldsymbol{\beta}} = (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{y}, \tag{8.4}$$

Substitution yields

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'(\,\mathbf{X}\boldsymbol{\beta} + \mathbf{u}\,) \\ &= (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{u}.\end{aligned} \tag{8.5}$$

8.1.3 Properties of $\widehat{\boldsymbol{\beta}}$

Since $\mathbf{X}$ is nonstochastic,

$$\begin{aligned}\mathrm{E}[\,\widehat{\boldsymbol{\beta}}\,] &= \boldsymbol{\beta} + \mathrm{E}[(\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{u}\,] \\ &= \boldsymbol{\beta} + (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\,\mathrm{E}[\,\mathbf{u}\,] \\ &= \boldsymbol{\beta}.\end{aligned} \tag{8.6}$$

Thus, the OLS estimator is unbiased. Also,

$$\begin{aligned}\mathrm{Cov}(\,\widehat{\boldsymbol{\beta}}\,) &= \mathrm{E}(\,\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\,)(\,\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\,)' \\ &= \mathrm{E}[(\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\,\mathbf{X}'\mathbf{X}\,)^{-1}] \\ &= (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\,\mathrm{E}(\,\mathbf{u}\mathbf{u}'\,)\mathbf{X}(\,\mathbf{X}'\mathbf{X}\,)^{-1} \\ &= (\,\mathbf{X}'\mathbf{X}\,)^{-1}\mathbf{X}'\sigma^2\mathrm{I}_n\mathbf{X}(\,\mathbf{X}'\mathbf{X}\,)^{-1} \\ &= \sigma^2(\,\mathbf{X}'\mathbf{X}\,)^{-1}.\end{aligned} \tag{8.7}$$

The elements of $\mathbf{X}'\mathbf{X}$ are unbounded as n gets very large. Therefore,

$$\lim_{n\to\infty}(\,\mathbf{X}'\mathbf{X}\,)^{-1} = 0, \tag{8.8}$$

and the variances of $\widehat{\boldsymbol{\beta}}$ converge to zero. This means that the distribution collapses about its expected value, namely $\boldsymbol{\beta}$. So,

$$\operatorname*{plim}_{n\to\infty}\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}, \tag{8.9}$$

and OLS estimation is consistent.

The OLS estimates $\widehat{\boldsymbol{\beta}}$ are the best linear unbiased (BLUE) in that they have minimum variance in the class of unbiased estimators of $\boldsymbol{\beta}$ that are also linear in $\mathbf{y}$.

Suppose that $\mathbf{u}$ is normal, then the linear transformation

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \tag{8.10}$$

is also normal.

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim \mathrm{N}(0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{8.11}$$

or

$$\widehat{\boldsymbol{\beta}} \sim \mathrm{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{8.12}$$

Moreover, the $\widehat{\boldsymbol{\beta}}$ are maximum likelihood and hence minimum variance in the class of unbiased estimators.

8.1.4 Properties of e

Now, the OLS residuals are

$$\begin{aligned}
\mathbf{e} &= \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \\
&= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= [\mathrm{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \\
&= \mathbf{M}\mathbf{y} \qquad \mathbf{M} = \mathrm{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \qquad (8.13) \\
&= \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\
&= [\mathrm{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'](\mathbf{X}\boldsymbol{\beta}) + \mathbf{M}\mathbf{u} \\
&= \mathbf{M}\mathbf{u}. \qquad (8.14)
\end{aligned}$$

Since $\mathbf{MX} = 0$. Thus, the OLS residuals are a linear transformation of the underlying disturbances. Also,

$$\begin{aligned}
\mathbf{X}'\mathbf{e} &= \mathbf{X}'\mathbf{M}\mathbf{u} \\
&= 0, \qquad (8.15)
\end{aligned}$$

again, since $\mathbf{MX} = 0$, and the OLS residuals are orthogonal or linearly unrelated to $\mathbf{X}$. When $\mathbf{u}$ are normal, then the linear transformation $\mathbf{e} = \mathbf{Mu}$ is also normal. Specifically,

$$\mathbf{e} \sim \mathrm{N}(0, \sigma^2\mathbf{M}) \tag{8.16}$$

since

$$\mathrm{E}\,\mathbf{e} = \mathrm{E}\,\mathbf{M}\mathbf{u} = \mathbf{M}\,\mathrm{E}\,\mathbf{u} = 0, \tag{8.17}$$

and

$$\begin{aligned}
\mathrm{E}\,\mathbf{e}\mathbf{e}' &= \mathrm{E}\,\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{M}' \\
&= \mathbf{M}(\mathrm{E}\,\mathbf{u}\mathbf{u}')\mathbf{M}' \\
&= \mathbf{M}\left(\sigma^2\mathrm{I}\right)\mathbf{M}' \\
&= \sigma^2\mathbf{M}, \qquad (8.18)
\end{aligned}$$

since $\mathbf{M}\mathbf{M}' = \mathbf{M}$.

8.2 Tests Based on the χ^2 Distribution

8.2.1 The χ^2 Distribution

Suppose that $z_1, z_2, \ldots, z_n$ are iid $\mathrm{N}(0,1)$ random variables. Then,

$$\sum_{i=1}^{n} z_i^2 \sim \chi_n^2. \tag{8.19}$$

8.2.2 Distribution of $(n-k)s^2/\sigma^2$

Now,

$$u_i = y_i - \mathbf{x}_i'\boldsymbol{\beta} \sim \mathrm{N}(0, \sigma^2), \tag{8.20}$$

so

$$\frac{u_i}{\sigma} \sim \mathrm{N}(0,1) \tag{8.21}$$

and

$$\sum_{i=1}^{n} \left(\frac{u_i}{\sigma}\right)^2 = \sum_{i=1}^{n} \frac{u_i^2}{\sigma^2} \sim \chi_n^2. \tag{8.22}$$

Now,

$$e_i = y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}} \tag{8.23}$$

is an estimate of u_i and we might expect that

$$\sum_{i=1}^{n} \frac{e_i^2}{\sigma^2} \sim \chi_n^2. \tag{8.24}$$

However, this would be wrong as only $n-k$ of the observations are independent since $\mathbf{e}$ satisfies the k equations $\mathbf{X}'\mathbf{e} = 0$.

The properties of $\mathbf{e} = \mathbf{M}\mathbf{u}$ follow from the properties of $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, which is symmetric idempotent and positive semi-definite and hence has some very special properties. First, $\mathrm{rank}(\mathbf{M}) = \mathrm{tr}(\mathbf{M}) = n-k$. Second, we can write the decomposition $\mathbf{M} = \mathbf{Q}\mathbf{D}_{n-k}\mathbf{Q}'$ where $\mathbf{D}_{n-k}$ is a diagonal matrix with its first $n-k$ diagonals unity and the remainder zero, and $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_n$ so $\mathbf{Q}' = \mathbf{Q}^{-1}$. Let

$$\mathbf{v} = \mathbf{Q}'\mathbf{u} \tag{8.25}$$

then $\mathbf{v} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}_n)$ and $\mathbf{u} = \mathbf{Q}\mathbf{v}$. Substitution yields

$$\begin{aligned} \frac{1}{\sigma^2}\mathbf{e}'\mathbf{e} &= \frac{1}{\sigma^2}\mathbf{u}'\mathbf{M}\mathbf{u} \qquad (8.26) \\ &= \frac{1}{\sigma^2}\mathbf{u}'\mathbf{Q}\mathbf{D}_{n-k}\mathbf{Q}'\mathbf{u} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sigma^2}\mathbf{v}'\mathbf{Q}'\mathbf{Q}\mathbf{D}_{n-k}\mathbf{Q}'\mathbf{Q}\mathbf{v} \\
&= \frac{1}{\sigma^2}\mathbf{v}'\mathbf{D}_{n-k}\mathbf{v} \\
&= \frac{1}{\sigma^2}\textstyle\sum_{i=1}^{n-k} v_i^2 \\
&= \textstyle\sum_{i=1}^{n-k} (\frac{v_i}{\sigma})^2 \sim \chi^2_{n-k}
\end{aligned}
$$

Thus,

$$\sum_{i=1}^{n} \frac{e_i^2}{\sigma^2} \sim \chi^2_{n-k} \tag{8.27}$$

and

$$(n-k)\frac{\frac{\sum_{i=1}^n e_i^2}{n-k}}{\sigma^2} = (n-k)\frac{s^2}{\sigma^2} \sim \chi^2_{n-k}. \tag{8.28}$$

Not only so but $\mathbf{e} = \mathbf{M}\mathbf{u}$ and $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ are jointly normal and

$$
\begin{aligned}
\mathrm{E}[\mathbf{e}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] &= \mathrm{E}[\mathbf{M}\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \qquad (8.29) \\
&= \mathbf{M}\sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2\mathbf{M}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = 0
\end{aligned}
$$

so they are uncorrelated and independent and s^2 is independent of $\widehat{\boldsymbol{\beta}}$, since it is a function only of $\mathbf{e}$.

8.2.3 A Confidence Interval

Now, let a and b be numbers such that

$$\Pr(b \le \chi^2_{n-k} \le a) = 0.95, \tag{8.30}$$

say. Then a and b can be obtained from a table. Thus,

$$
\begin{aligned}
\Pr\left(b \le (n-k)\frac{s^2}{\sigma^2} \le a\right) &= 0.95 \\
\Pr\left(\frac{1}{b} \ge \frac{\sigma^2}{(n-k)s^2} \ge \frac{1}{a}\right) &= 0.95 \\
\Pr\left(\frac{(n-k)s^2}{b} \ge \sigma^2 \ge \frac{(n-k)s^2}{a}\right) &= 0.95 \qquad (8.31)
\end{aligned}
$$

establishes a 95% confidence interval for σ^2.

For example, for $n - k = 14$, we have

$$\Pr\left(5.64 \leq \frac{14s^2}{\sigma^2} \leq 26.12\right) = 0.95$$

$$\Pr\left(\frac{14s^2}{5.63} \geq \sigma^2 \geq \frac{14s^2}{26.12}\right) = 0.95 \qquad (8.32)$$

and if $s^2 = 4.0$, then

$$\Pr\left(\frac{56}{5.63} \geq \sigma^2 \geq \frac{56}{26.12}\right) = 0.95 \qquad (8.33)$$

or

$$\Pr\left(10 \geq \sigma^2 \geq 2.1\right) = 0.95 \qquad (8.34)$$

is the confidence interval.

8.2.4 A Hypothesis Test

Suppose that

$$\mathrm{H}_0\colon \sigma^2 = \sigma_0^2, \qquad \mathrm{H}_1\colon \sigma^2 \neq \sigma_0^2.$$

Then we know that

$$(n-k)\frac{s^2}{\sigma_0^2} \sim \chi^2_{n-k}. \qquad (8.35)$$

under the null hypothesis.

Choose $\alpha = 0.05$, say, then critical values corresponding to 2.5% tails are 5.63 and 26.12 for $n - k = 14$. Thus, if

$$5.62 \leq 14\frac{s^2}{\sigma_0^2} \leq 26.12, \qquad (8.36)$$

we fail to reject the null hypothesis. Otherwise, we reject it at the 5% level of confidence. For example, suppose that $s^2 = 4.0$ and $\sigma_0^2 = 1$, then

$$14\frac{s^2}{\sigma_0^2} = 56 \qquad (8.37)$$

and we reject the null hypothesis since we fall into the right-hand 2.5% tail.

8.3 Tests Based on the t Distribution

8.3.1 The t Distribution

Suppose that z is a $\mathrm{N}(0,1)$ random variable and that $w \sim \chi^2_m$ independent of z. Then,

$$\frac{z}{\sqrt{\frac{w}{m}}} \sim t_m. \qquad (8.38)$$

8.3.2 The Distribution of $(\widehat{\beta}_i - \beta_i)/(s^2 d_{ii})^{1/2}$

We have seen that

$$\widehat{\beta}_i \sim \mathrm{N}(\beta_i, \sigma^2 d_{ii}), \tag{8.39}$$

where d_{ii} is the (i, i) element of the matrix $(\mathbf{X}'\mathbf{X})^{1/2}$. Then,

$$z = \frac{\widehat{\beta}_i - \beta_i}{\sqrt{\sigma^2 d_{ii}}} \sim \mathrm{N}(1, 0), \tag{8.40}$$

while

$$w = (n-k)\frac{s^2}{\sigma_0^2} \sim \chi^2_{n-k}. \tag{8.41}$$

Since $\widehat{\beta}$ and s^2 are independent, we have

$$\frac{\frac{\widehat{\beta}_i - \beta_i}{\sqrt{\sigma^2 d_{ii}}}}{\sqrt{\frac{(n-k)s^2}{\sigma_0^2}/(n-k)}} = \frac{\widehat{\beta}_i - \beta_i}{\sqrt{s^2 d_{ii}}} \sim t_{n-k}. \tag{8.42}$$

8.3.3 Confidence Interval for β_i

First, obtain a such that

$$\Pr(-a \le t_{n-k} \le a) = 0.95, \tag{8.43}$$

say, from a table. Then,

$$\begin{aligned} \Pr\left(-a \le \frac{\widehat{\beta}_i - \beta_i}{\sqrt{\sigma^2 d_{ii}}} \le a\right) &= 0.95 \\ \Pr\left(\widehat{\beta}_i + a\sqrt{\sigma^2 d_{ii}} \ge \frac{\widehat{\beta}_i - \beta_i}{\sqrt{\sigma^2 d_{ii}}} \ge \widehat{\beta}_i - a\sqrt{\sigma^2 d_{ii}}\right) &= 0.95 \end{aligned} \tag{8.44}$$

and $\widehat{\beta}_i \pm a\sqrt{\sigma^2 d_{ii}}$ defines a 95% confidence interval for β_i.

8.3.4 Testing a Hypothesis

Suppose that

$$\mathrm{H}_0\colon \beta_i = \beta_i^0, \qquad \mathrm{H}_1\colon \beta_i \ne \beta_i^0.$$

We know that

$$\frac{\widehat{\beta}_i - \beta_i}{\sqrt{\sigma^2 d_{ii}}} \sim t_{n-k}. \tag{8.45}$$

Now, choose $\alpha = 0.05$, say, then critical values corresponding to 2.5% tails of a t distribution with $(n-k)$ degrees of freedom are $\pm a$, say, so if

$$-a \leq \frac{\widehat{\beta}_i - \beta_i}{\sqrt{\sigma^2 d_{ii}}} \leq a, \tag{8.46}$$

we fail to reject the null hypothesis. Otherwise we reject the null hypothesis in favor of the alternative.

8.3.5 The Distribution of $\mathbf{c}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/(s^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2}$

Consider the linear combination

$$\mathbf{c}'\widehat{\boldsymbol{\beta}}. \tag{8.47}$$

Then,

$$\mathbf{c}'\widehat{\beta} \sim \mathrm{N}(\mathbf{c}'\boldsymbol{\beta}, \sigma^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}), \tag{8.48}$$

Then,

$$\frac{\mathbf{c}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{\sigma^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}} \sim \mathrm{N}(1, 0). \tag{8.49}$$

As before, we use s^2 instead of σ^2, so while

$$\frac{\mathbf{c}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sqrt{s^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}} \sim t_{n-k}. \tag{8.50}$$

We can perform inferences and calculate confidence intervals as before.

8.4 Tests Based on the F Distribution

8.4.1 The F Distribution

Suppose that

$$v \sim \chi^2_l \quad \text{and} \quad v \sim \chi^2_m$$

If v and w are independent, then

$$\frac{v/l}{w/m} \sim \mathrm{F}_{l,m}. \tag{8.51}$$

8.4.2 Distribution of $(R\widehat{\beta} - r)'[s^2R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)/q$

Suppose we are interested in testing a set of q linear restrictions. Examples would be $\beta_1 + \beta_2 + ... + \beta_k = 1$ and $\beta_3 = 2\beta_2$. More generally, we consider

$$\mathrm{H}_0\colon\ \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \qquad \mathrm{H}_1\colon \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$$

where $\mathbf{r}$ is a $q \times 1$ known vector and $\mathbf{R}$ is a $q \times k$ known matrix. Due to the multivariate normality of $\widehat{\boldsymbol{\beta}}$, then under the null hypothesis, we have

$$\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r} \sim N(\mathbf{0}, \sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}') \tag{8.52}$$

and hence

$$(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'[\sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')]^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}) \sim \chi^2_q. \tag{8.53}$$

Recall that $\widehat{\boldsymbol{\beta}}$ and s^2 are independent so $(n-k)\frac{s^2}{\sigma^2} \sim \chi^2_{n-k}$ is independent of the quadratic form in (8.53). Thus, under the null hypothesis,

$$\frac{(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'[\sigma^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')]^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})/\mathbf{q}}{(n-k)\frac{s^2}{\sigma^2}/(n-k)} \sim F_{q,n-k} \tag{8.54}$$

and after some simplification

$$(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'[s^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')]^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})/\mathbf{q} \sim F_{q,n-k}. \tag{8.55}$$

Under the alternative hypothesis $\mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$, then the numerator diverges at the rate n and we expect large positive values of the statistic with high probability. Accordingly, we only consult the RHS tail values of the distribution to establish critical values. Values of the statistic exceeding these critical values are rare events under the null but typical under the alternative, so we reject when the realization exceeds the critical value.

8.4.3 The Distribution of $[(\mathrm{SSE}_r - \mathrm{SSE}_u)/q]/\mathrm{SSE}_u/(n-k)$

The most common form of linear restrictions that occur are zero restrictions. Suppose the model of interest can be written as

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u} \tag{8.56}$$

and

$$\mathrm{H}_0\colon \boldsymbol{\beta}_2 = 0 \qquad \mathrm{H}_1\colon \boldsymbol{\beta}_2 \neq 0.$$

Define the "unrestricted" residuals

$$\mathbf{e}_u = \mathbf{y} - \mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 \tag{8.57}$$

and

$$\mathrm{SSE}_u = \mathbf{e}_u'\mathbf{e}_u \tag{8.58}$$

from the OLS regression of $\mathbf{y}$ on $\mathbf{X}_1$ and $\mathbf{X}_2$.

Next, define the "restricted" residuals

$$\mathbf{e}_r = \mathbf{y} - \mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 \tag{8.59}$$

and

$$\mathrm{SSE}_r = \mathbf{e}_r'\mathbf{e}_r \tag{8.60}$$

from the OLS regression of $\mathbf{y}$ on $\mathbf{X}_1$ only.

Now, $\mathrm{SSE}_r \geq \mathrm{SSE}_u$, but under $\mathrm{H}_0\colon \beta_2 = 0$, we expect

$$\frac{\mathrm{SSE}_u}{\sigma^2} \sim \chi^2_{n-(k_1+k_2)} \qquad \text{and} \qquad \frac{\mathrm{SSE}_r}{\sigma^2} \sim \chi^2_{n-k_1} \tag{8.61}$$

to have similar values. We therefore might expect

$$\frac{\mathrm{SSE}_r/(n-k_1)}{\mathrm{SSE}_u/(n-(k_1+k_2))} \sim \mathrm{F}_{n-k_1,n-(k_1+k_2)}, \tag{8.62}$$

but unfortunately, SSE_r and SSE_u are not independent because they both satisfy

$$\mathbf{X}_1'\mathbf{e}_u = \mathbf{X}_1'\mathbf{e}_r = 0. \tag{8.63}$$

The appropriate ratio can be determined by applying the results of the previous section. Specifically, we take $\mathbf{R} = (\mathbf{0} : \mathbf{I}_{k_2})$ and $\mathbf{r} = \mathbf{0}$ whereupon the restictions $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ are equivalent to $\boldsymbol{\beta}_2 = \mathbf{0}$. For this choice of $\mathbf{R}$ and $\mathbf{r}$ we have $\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r} = \widehat{\boldsymbol{\beta}}_2$ and using the results for inverses of partitioned matrices

$$\begin{aligned} \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' &= (\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2)^{-1} \\ &= (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1} \end{aligned} \tag{8.64}$$

where $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$. Substitution yields

$$\begin{aligned} (\mathbf{R}\widehat{\boldsymbol{\beta}}-\mathbf{r})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')]^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}}-\mathbf{r})/\mathbf{q} &= \widehat{\boldsymbol{\beta}}_2'[\sigma^2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}]^{-1}\widehat{\boldsymbol{\beta}}_2 \\ &= \frac{1}{\sigma^2}\widehat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 \\ &= \frac{1}{\sigma^2}\widehat{\boldsymbol{\beta}}_2'\mathbf{X}_2'\mathbf{M}_1\mathbf{M}_1\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 \\ &= \frac{1}{\sigma^2}(\mathbf{M}_1\mathbf{y} - \mathbf{e}_u)'(\mathbf{M}_1\mathbf{y} - \mathbf{e}_u) \\ &= \frac{1}{\sigma^2}(\mathbf{y}'\mathbf{M}_1\mathbf{y} - \mathbf{2}\mathbf{e}_u'\mathbf{M}_1\mathbf{y} - \mathbf{e}_u'\mathbf{e}_u) \\ &= \frac{1}{\sigma^2}(\mathbf{y}'\mathbf{M}_1\mathbf{y} - \mathbf{e}_u'\mathbf{e}_u) \\ &= \frac{1}{\sigma^2}(\mathbf{e}_r'\mathbf{e}_r - \mathbf{e}_u'\mathbf{e}_u) \end{aligned} \tag{8.65}$$

where we use the results

$$\begin{aligned} \mathbf{M}_1\mathbf{y} &= \mathbf{M}_1(\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 + \mathbf{e}_u) \\ &= \mathbf{M}_1(\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 + \mathbf{e}_u) \\ &= \mathbf{M}_1\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 + \mathbf{M}_1\mathbf{e}_u \end{aligned} \tag{8.66}$$

and $\mathbf{M}_1\mathbf{e}_u = \mathbf{M}_1\mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{y} = \mathbf{e}_u$.

Since $\mathrm{SSE}_r = \mathbf{e}_r'\mathbf{e}_r$ and $\mathrm{SSE}_u = \mathbf{e}_u'\mathbf{e}_u$, then $\mathrm{SSE}_r - \mathrm{SSE}_u$ is the sum-of-squares with k_2 degrees of freedom that are independent of SSE_u and form (canceling σ^2 in the numerator and denominator)

$$\frac{(\mathrm{SSE}_r - \mathrm{SSE}_u)/k_2}{\mathrm{SSE}_u/(n-(k_1+k_2))} \sim \mathrm{F}_{k_1, n-(k_1+k_2)}. \tag{8.67}$$

Under the null hypothesis, this value will usually be small. Under the alternative of $\beta_2 \neq 0$, however, we would expect SSE_u to be much smaller than SSE_r and the above ratio to be large.

8.4.4 Testing a Hypothesis

We can consult the tables to find the critical point, c, corresponding to $\alpha = 0.05$, say. Then, if

$$\frac{(\mathrm{SSE}_r - \mathrm{SSE}_u)/k_1}{\mathrm{SSE}_u/(n-(k_1+k_2))} > c, \tag{8.68}$$

we reject the null hypothesis at the 5% level.

Note that for $k_1 = 1$, that is, one restriction,

$$\sqrt{\frac{(\mathrm{SSE}_r - \mathrm{SSE}_u)/k_1}{\mathrm{SSE}_u/(n-(k_1+k_2))}} \sim t_{n-(k_1+k_2)} \tag{8.69}$$

8.5 An Example

Consider the model

$$Y_t = \beta_1 + \beta_2 X_{t,2} + \beta_3 X_{t,3} + u_t, \tag{8.70}$$

where Y_t is wheat yield, $X_{t,2}$ is the amount of fertilizer applied and $X_{t,3}$ is the annual rainfall. The data are given in Table 8.5. After we rescale the data, we obtain the estimates of the βs given in Table 8.5.

Now,

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 55.9141 & -0.4189 & -15.4560 \\ -0.4189 & 0.0371 & 0.0773 \\ -15.4560 & 0.0773 & 4.3277 \end{pmatrix}, \tag{8.71}$$

and $s^2 = \sum_t e_t^2/(T-3) = 0.5232/4 = 0.1308$. Now, $\mathrm{R}^2 = 1 - 0.5232/13.5 = 0.9612$ and $\overline{\mathrm{R}}^2 = 1 - 6/4 \cdot 0.0388 = 0.9419$. Recall

$$\mathrm{E}[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}, \tag{8.72}$$

which we estimate using

$$\text{est. } \mathrm{Cov}(\widehat{\beta}) = s^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{8.73}$$

Wheat Yield (Bushels/Acre)	Fertilizer (Pounds/Acre)	Rainfall (Inches/Year)
40	100	36
45	200	33
50	300	37
65	400	37
70	500	34
70	600	32
80	700	36

Table 8.1: Wheat yield data.

Parameter	Estimate
β_1	1.1329
β_2	0.6893
β_3	0.6028

Table 8.2: Wheat yield parameter estimates.

Thus,

$$\begin{aligned}\text{est. Var}(\widehat{\beta}_1) &= s^2 d_{11} = 7.3133 \\ \text{est. Var}(\widehat{\beta}_2) &= s^2 d_{22} = 0.0049 \\ \text{est. Var}(\widehat{\beta}_3) &= s^2 d_{33} = 0.5660\end{aligned}$$

For $\text{H}_0\colon \beta_1 = 0$ vs $\text{H}_1\colon \beta_1 \neq 0$, we have

$$\frac{\widehat{\beta}_1 - \beta_1^0}{\sqrt{\text{Var}(\widehat{\beta}_1)}} = \frac{1.1329}{2.7043} = 0.4189 \sim t_4. \tag{8.74}$$

Now, a 95% acceptance region for a t_4 distribution is $-2.776 \leq t_4 \leq 2.776$. Thus, we fail to reject the null hypothesis.

For $\text{H}_0\colon \beta_2 = 0$ vs $\text{H}_1\colon \beta_2 \neq 0$, we have

$$\frac{\widehat{\beta}_2 - \beta_2^0}{\sqrt{\text{Var}(\widehat{\beta}_2)}} = \frac{0.6893}{0.0697} = 9.8965 \sim t_4. \tag{8.75}$$

and we reject the null hypothesis at the 95% confidence level. In fact, we reject at the 99.9% confidence level, where the acceptance region is $-7.173 \leq t_4 \leq 7.173$.