CHAPTER III

# The Univariate Normal Linear Regression Model

In this chapter we first take up the analysis of the simple univariate normal linear regression model and then turn to the normal linear multiple regression model.[1] Throughout we adopt assumptions of normality, independence, linearity, homoscedasticity, and absence of measurement errors. Certain departures from these specifying assumptions and their analysis are treated in subsequent chapters.

## 3.1 THE SIMPLE UNIVARIATE NORMAL LINEAR REGRESSION MODEL

### 3.1.1 Model and Likelihood Function

In the simple univariate normal linear regression model we have *one* random variable (hence the term "univariate"), the "dependent" variable, whose variation is to be explained, at least in part, by the variation of another variable, the "independent" variable. That part of the variation of the dependent variable unexplained by variation in the independent variable is assumed to be produced by an unobserved random "error" or "disturbance" variable which may be viewed as representing the collective action of a number of minor factors that produce variation in the dependent variable. Formally, with the dependent variable, denoted by $y$, and the independent variable denoted by $x$, we have the following relationship:

$$y_i = \beta_1 + \beta_2 x_i + u_i, \qquad i = 1, 2, \ldots, n, \tag{3.1}$$

[1] Bayesian analyses of the univariate normal linear regression model appear in H. Jeffreys, *Theory of Probability* (3rd rev. ed.). Oxford: Clarendon, 1966, pp. 147–161; D. V. Lindley, *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 2. Inference.* Cambridge: Cambridge University Press, 1965, Chapter 8; and H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory.* Boston: Graduate School of Business Administration, Harvard University, 1961, Chapter 13.





where $y_i$ = $i$th observation on the dependent variable,
$x_i$ = $i$th observation on the independent variable,
$u_i$ = $i$th unobserved value of the random disturbance or error variable, and
$\beta_1$ and $\beta_2$ = regression parameters, namely, the "intercept" and "slope coefficient," respectively.

Note that the relation in (3.1) is linear in $\beta_1$, $\beta_2$, and $u_i$, hence the term "linear" regression.[2]

**Assumption 1.** The $u_i$, $i = 1, 2, \ldots, n$, are normally and independently distributed, each with zero mean and common variance $\sigma^2$.

Regarding the independent variable, we make the following assumption:

**Assumption 2.** The $x_i$, $i = 1, 2, \ldots, n$, are fixed nonstochastic variables.

Alternatively, we can make the following assumption about $x_i$:

**Assumption 3.** The $x_i$, $i = 1, 2, \ldots, n$, are random variables distributed independently of the $u_i$, with a pdf *not* involving the parameters $\beta_1$, $\beta_2$, and $\sigma$.

To form the likelihood function under assumptions 1 and 3, we write the joint pdf for $\mathbf{y}' = (y_1, y_2, \ldots, y_n)$ and $\mathbf{x}' = (x_1, x_2, \ldots, x_n)$, namely

$$p(\mathbf{y}, \mathbf{x}|\beta_1, \beta_2, \sigma^2, \boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{x}, \beta_1, \beta_2, \sigma^2)\, g(\mathbf{x}|\boldsymbol{\theta}), \tag{3.2}$$

when $\boldsymbol{\theta}$ denotes the parameters of the marginal pdf for $\mathbf{x}$. Since by assumption (3) $\boldsymbol{\theta}$ does not involve $\beta_1$, $\beta_2$, or $\sigma$, the likelihood function for $\beta_1$, $\beta_2$, and $\sigma$ can be formed from the first factor on the right-hand side of (3.2). Note from (3.1) that for given $\mathbf{x}$, $\beta_1$, $\beta_2$, and $\sigma^2$, $\mathbf{y}$ will be normally distributed with $E(y_i|x_i, \beta_1, \beta_2, \sigma^2) = \beta_1 + \beta_2 x_i$ and $\mathrm{Var}(y_i|x_i, \beta_1, \beta_2, \sigma^2) = \sigma^2$, $i = 1, 2, \ldots, n$. Further, the $y_i$, given the $x_i$, $\beta_1$, $\beta_2$, and $\sigma^2$, will be independently distributed. Thus we have

$$p(\mathbf{y}|\mathbf{x}, \beta_1, \beta_2, \sigma) \propto \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2} \sum (y_i - \beta_1 - \beta_2 x_i)^2\right], \tag{3.3}$$

with the summation extending from $i = 1$ to $i = n$. Also, (3.3) would have resulted had we adopted Assumption 2 about the $x_i$ rather than Assumption 3. The expression in (3.3), viewed as a function of the parameters $\beta_1$, $\beta_2$, and $\sigma$, is the likelihood function to be combined with our prior pdf for the parameters.

[2] The relation in (3.1) need not be linear in the "underlying" variables; for example, it may be that $y_i = \log w_i$, where $w_i$, is the $i$th observation on an underlying variable, or $x_i$ may represent $z_i^2$, where $z_i$ is an observation on an underlying variable.

### 3.1.2 Posterior Pdf's for Parameters with a Diffuse Prior Pdf

For our prior pdf for $\beta_1$, $\beta_2$ and $\sigma$ we assume that $\beta_1$, $\beta_2$ and $\log \sigma$ are uniformly and independently distributed, which implies

$$p(\beta_1, \beta_2, \sigma) \propto \frac{1}{\sigma} \qquad \begin{matrix} -\infty < \beta_1, \beta_2 < \infty, \\ 0 < \sigma < \infty. \end{matrix} \tag{3.4}$$

Then, on combining (3.3) and (3.4), the joint posterior pdf for $\beta_1$, $\beta_2$, and $\sigma$ is given by

$$p(\beta_1, \beta_2, \sigma | \mathbf{y}, \mathbf{x}) \propto \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2} \sum (y_i - \beta_1 - \beta_2 x_i)^2\right]. \tag{3.5}$$

This joint posterior pdf, which serves as a basis for making inferences about $\beta_1$, $\beta_2$, and $\sigma$, can be analyzed conveniently by taking note of the following algebraic identity[3]:

$$\begin{aligned} \sum (y_i - \beta_1 - \beta_2 x_i)^2 &= \nu s^2 + n(\beta_1 - \hat{\beta}_1)^2 \\ &\quad + (\beta_2 - \hat{\beta}_2)^2 \sum x_i^2 + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) \sum x_i, \end{aligned} \tag{3.6}$$

where $\nu = n - 2$,

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}, \qquad \hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}, \tag{3.7}$$

$$s^2 = \nu^{-1} \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2, \tag{3.8}$$

with $\bar{y} = n^{-1} \sum y_i$ and $\bar{x} = n^{-1} \sum x_i$. To establish (3.6) we write

$$\sum (y_i - \beta_1 - \beta_2 x_i)^2 = \sum \{(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) - [(\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2) x_i]\}^2.$$

On expanding the rhs, note that the cross-product term vanishes and thus (3.6) results.

On substituting from (3.6) in (3.5), we have

$$\begin{aligned} p(\beta_1, \beta_2, \sigma | \mathbf{y}, \mathbf{x}) &\propto \frac{1}{\sigma^{n+1}} \\ &\times \exp\left\{-\frac{1}{2\sigma^2} [\nu s^2 + n(\beta_1 - \hat{\beta}_1)^2 + (\beta_2 - \hat{\beta}_2)^2 \sum x_i^2 \right. \\ &\qquad \left. + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) \sum x_i]\right\}. \end{aligned} \tag{3.9}$$

From (3.9) it is immediately seen that the conditional posterior pdf for $\beta_1$ and $\beta_2$, given $\sigma$, is in the bivariate normal form with mean $(\hat{\beta}_1, \hat{\beta}_2)$ and

[3] If the expression in (3.6) is substituted in (3.3), the likelihood function can be expressed in terms of $s^2$, $\hat{\beta}_1$, and $\hat{\beta}_2$, which are sufficient statistics.



covariance matrix

$$\sigma^2 \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}^{-1} = \sigma^2 \begin{bmatrix} \dfrac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} & \dfrac{-\bar{x}}{\sum (x_i - \bar{x})^2} \\ \dfrac{-\bar{x}}{\sum (x_i - \bar{x})^2} & \dfrac{1}{\sum (x_i - \bar{x})^2} \end{bmatrix}.$$

Of course, since $\sigma^2$ is rarely known in practice, this result is not very useful. To obtain the marginal posterior pdf for $\beta_1$ and $\beta_2$ we integrate (3.9) with respect to $\sigma$ to obtain

$$\begin{aligned} p(\beta_1, \beta_2 | \mathbf{y}, \mathbf{x}) &= \int_0^\infty p(\beta_1, \beta_2, \sigma | \mathbf{y}, \mathbf{x})\, d\sigma \\ &\propto [\nu s^2 + n(\beta_1 - \hat{\beta}_1)^2 + (\beta_2 - \hat{\beta}_2)^2 \sum x_i^2 \\ &\qquad + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) \sum x_i]^{-n/2}, \end{aligned} \tag{3.10}$$

which is seen to be in the form of a bivariate Student $t$ pdf (see Appendix B). From properties of the bivariate Student $t$ pdf we have the following results:

$$p(\beta_1 | \mathbf{y}, \mathbf{x}) \propto \left[\nu + \frac{\sum (x_i - \bar{x})^2}{s^2 \sum x_i^2 / n} (\beta_1 - \hat{\beta}_1)^2\right]^{-(\nu+1)/2}, \qquad -\infty < \beta_1 < \infty, \tag{3.11}$$

$$p(\beta_2 | \mathbf{y}, \mathbf{x}) \propto \left[\nu + \frac{\sum (x_i - \bar{x})^2}{s^2} (\beta_2 - \hat{\beta}_2)^2\right]^{-(\nu+1)/2}, \qquad -\infty < \beta_2 < \infty. \tag{3.12}$$

If we make the following transformations in (3.11) and (3.12),

$$\left[\frac{\sum (x_i - \bar{x})^2}{s^2 \sum x_i^2 / n}\right]^{1/2} (\beta_1 - \hat{\beta}_1) = t_\nu \tag{3.13}$$

and

$$\frac{\beta_2 - \hat{\beta}_2}{s / [\sum (x_i - \bar{x})^2]^{1/2}} = t_\nu, \tag{3.14}$$

the random variable $t_\nu$ has the Student $t$ pdf with $\nu$ degrees of freedom. These results enable us to make inferences about $\beta_1$ and $\beta_2$, using tables of the $t$-distribution.

As regards the posterior pdf for $\sigma$, it can be obtained by integrating (3.9) with respect to $\beta_1$ and $\beta_2$. This operation yields

$$p(\sigma | \mathbf{y}, \mathbf{x}) \propto \frac{1}{\sigma^{\nu+1}} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right), \qquad 0 < \sigma < \infty. \tag{3.15}$$

From (3.15) $\sigma$ is distributed in the form of an inverted gamma function (see Appendix A). Thus we have[4]

$$E\sigma = s\left(\sqrt{\frac{\nu}{2}}\right)^{1/2} \frac{\Gamma[(\nu - 1)/2]}{\Gamma(\nu/2)}; \qquad \operatorname{Var}(\sigma) = \frac{s^2 \nu}{\nu - 2} - (E\sigma)^2.$$

[4] For the pdf in (3.15) to be proper we must have $\nu > 0$; for the mean to exist we need $\nu > 1$; and for the variance to exist we need $\nu > 2$. See Appendix A.

Further, if we transform from $\sigma$ to $\sigma^2$, the posterior pdf for the variance is

$$(3.16) \qquad p(\sigma^2|\mathbf{y}, \mathbf{x}) \propto [(\sigma^2)^{n/2}]^{-1} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right), \qquad 0 < \sigma^2 < \infty.$$

Finally, the posterior pdf for the precision parameter $h = 1/\sigma^2$ is given by

$$(3.17) \qquad p(h|\mathbf{y}, \mathbf{x}) \propto h^{\nu/2-1} \exp\left(-\frac{\nu s^2 h}{2}\right), \qquad 0 < h < \infty.$$

It is seen from (3.17) that the variable $\nu s^2 h$ has the $\chi^2$ pdf with $\nu$ degrees of freedom. From (3.16) and (3.17) we have, for example, $E\sigma^2 = \nu s^2/(\nu - 2)$ and $Eh = 1/s^2$. Further properties of these pdf's are discussed in Appendix A.

To make joint posterior inferences about $\beta_1$ and $\beta_2$ we shall show that the following quantity

$$(3.18) \qquad \psi = \frac{[n(\beta_1 - \hat\beta_1)^2 + (\beta_2 - \hat\beta_2)^2 \sum x_i^2 + 2(\beta_1 - \hat\beta_1)(\beta_2 - \hat\beta_2)\sum x_i]}{2s^2}$$

is distributed a posteriori as $F_{2,\nu}$. To show this let us write

$$\psi = \boldsymbol{\delta}' A \boldsymbol{\delta},$$

where $\boldsymbol{\delta}' = (\beta_1 - \hat\beta_1, \beta_2 - \hat\beta_2)$ and

$$(3.19) \qquad A \equiv \frac{1}{2s^2}\begin{bmatrix} 1 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}.$$

Using this notation, the posterior pdf for $\boldsymbol{\delta}$ is given from (3.10) by

$$(3.20) \qquad p(\boldsymbol{\delta}|\mathbf{y}, \mathbf{x}) \propto \left(1 + \frac{2}{\nu}\boldsymbol{\delta}' A \boldsymbol{\delta}\right)^{-n/2}$$

Now, since $A$ is positive definite, we can write $A = K'K$, where $K$ is non-singular and thus $\boldsymbol{\delta}' A \boldsymbol{\delta} = (K\boldsymbol{\delta})'K\boldsymbol{\delta} = \mathbf{V}'\mathbf{V}$, where $\mathbf{V} = K\boldsymbol{\delta}$ is a $2 \times 1$ vector and $\mathbf{V}' = (v_1, v_2)$. Then

$$(3.21) \qquad p(\mathbf{V}|\mathbf{y}, \mathbf{x}) \propto \left(1 + \frac{2}{\nu}\mathbf{V}'\mathbf{V}\right)^{-n/2}.$$

Now let

$$v_1 = \psi^{1/2} \cos\theta,$$
$$v_2 = \psi^{1/2} \sin\theta.$$

The Jacobian of this transformation is $\frac{1}{2}$. Note, too, that $\mathbf{V}'\mathbf{V} = v_1^2 + v_2^2 = \psi(\cos^2\theta + \sin^2\theta) = \psi$. Thus

$$(3.22) \qquad p(\psi|\mathbf{y}, \mathbf{x}) \propto \left(1 + \frac{2}{\nu}\psi\right)^{-(\nu+2)/2},$$



which is the $F_{2,\nu}$ pdf.[5] This result can be employed to construct posterior confidence regions for $\beta_1$ and $\beta_2$.

In (3.10) we noted that $\beta_1$ and $\beta_2$ are distributed in the bivariate Student $t$ form. An important property of this bivariate pdf is that a single linear combination of variables so distributed has a pdf in the univariate Student $t$ form. This result is illustrated below in obtaining the posterior pdf of $\eta_0$, defined by

$$(3.23) \qquad E(y|x = x_0) = \eta_0 = \beta_1 + \beta_2 x_0.$$

It is seen that $\eta_0$ is a linear form in $\beta_1$ and $\beta_2$ and thus will be distributed in the univariate Student $t$ form with mean $\hat\eta_0 = \hat\beta_1 + \hat\beta_2 x_0$; that is

$$(3.24) \qquad \frac{\eta_0 - \hat\eta_0}{s[1/n + (x_0 - \bar x)^2/\sum(x_i - \bar x)^2]^{1/2}} \sim t_\nu.$$

The result in (3.24) can be derived by changing variables in (3.10) from $\beta_1$ and $\beta_2$ to $\eta_0$ and $\beta_2$ as follows:

$$\eta_0 - \hat\eta_0 = \beta_1 - \hat\beta_1 + (\beta_2 - \hat\beta_2)x_0,$$
$$\beta_2 - \hat\beta_2 = \beta_1 - \hat\beta_2.$$

The Jacobian of this transformation is 1. Then, on integrating out $\beta_2$, the marginal posterior pdf for $\eta_0$ is given by

$$(3.25) \qquad p(\eta_0|\mathbf{y}, \mathbf{x}) \propto \left[\nu + \frac{n\sum(x_i - \bar x)^2}{s^2\sum(x_i - x_0)^2}(\eta_0 - \hat\eta_0)^2\right]^{-(\nu+1)/2}, \qquad -\infty < \eta_0 < \infty.$$

Then note that $\sum(x_i - x_0)^2 = \sum[(x_i - \bar x) - (x_0 - \bar x)]^2 = \sum(x_i - \bar x)^2 + n(x_0 - \bar x)^2$ and thus (3.24) follows. The result in (3.25) provides the complete posterior pdf for $\eta_0$ and (3.24) can be utilized to construct posterior intervals for $\eta_0$.

[5] In general $p(F) \propto F^{(m-2)/2}/(1 + m/qF)^{(m+q)/2}$ is the $F_{m,q}$ pdf with $0 < F < \infty$; see Appendix A.

## 3.2 THE NORMAL MULTIPLE REGRESSION MODEL

### 3.2.1 Model and Likelihood Function

With the normal multiple regression model, we assume that an $n \times 1$ vector of observations $\mathbf{y}$ on our dependent variable satisfies

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}, \tag{3.26}$$

where $X$ = an $n \times k$ matrix, with rank $k$, of observations on $k$ independent variables,

$\boldsymbol{\beta}$ = a $k \times 1$ vector of regression coefficients,

$\mathbf{u}$ = an $n \times 1$ vector of disturbance or error terms.

We assume that the elements of $\mathbf{u}$ are normally and independently distributed, each with mean zero and common variance $\sigma^2$; that is $E\mathbf{u} = \mathbf{0}$ and $E\mathbf{uu}' = \sigma^2 I_n$, where $I_n$ is an $n \times n$ unit matrix. With respect to the matrix $X$, if the regression equation is assumed to have a nonzero intercept, all elements in the first column of $X$ will be ones; that is, the first column is $\boldsymbol{\iota}$ with $\boldsymbol{\iota}' = (1, 1, \ldots, 1)$. The remaining elements of $X$ may be nonstochastic or stochastic, as in Section 3.1. If elements of $X$ are stochastic, it is assumed that they are distributed independently of $\mathbf{u}$ with a distribution that does *not* involve the parameters $\boldsymbol{\beta}$ and $\sigma$.

Under the above assumptions the joint pdf for the elements of $\mathbf{y}$, given $X$, $\boldsymbol{\beta}$, and $\sigma$, is

$$(3.27) \qquad \begin{aligned} p(\mathbf{y}|X, \boldsymbol{\beta}, \sigma) &\propto \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\beta})'(\mathbf{y} - X\boldsymbol{\beta})\right] \\ &\propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}[\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]\right\}, \end{aligned}$$

where $\nu = n - k$,

$$(3.28) \qquad \hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}.$$

and

$$(3.29) \qquad s^2 = \frac{(\mathbf{y} - X\hat{\boldsymbol{\beta}})'(\mathbf{y} - X\hat{\boldsymbol{\beta}})}{\nu}$$

are sufficient statistics. The second line of (3.27) makes use of the following algebraic identity:

$$\begin{aligned} (\mathbf{y} - X\boldsymbol{\beta})'(\mathbf{y} - X\boldsymbol{\beta}) &= [\mathbf{y} - X\hat{\boldsymbol{\beta}} - X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]'[\mathbf{y} - X\hat{\boldsymbol{\beta}} - X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})] \\ &= (\mathbf{y} - X\hat{\boldsymbol{\beta}})'(\mathbf{y} - X\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}), \end{aligned}$$

since the cross-product terms

$$(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'(\mathbf{y} - X\hat{\boldsymbol{\beta}}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'[X'\mathbf{y} - X'X(X'X)^{-1}X'\mathbf{y}] = 0.$$

### 3.2.2 Posterior Pdf's for Parameters with a Diffuse Prior Pdf

As prior pdf in the analysis of the multiple regression model, we assume that our information is diffuse or vague and represent it by taking the elements of $\boldsymbol{\beta}$ and $\log \sigma$ independently and uniformly distributed; that is

$$(3.30) \qquad p(\boldsymbol{\beta}, \sigma) \propto \frac{1}{\sigma}, \qquad \begin{aligned} -\infty < \beta_i < \infty, \\ 0 < \sigma < \infty, \end{aligned} \qquad \text{for } i = 1, 2, \ldots, k.$$

On combining (3.27) and (3.30), the joint posterior pdf for the parameters $\boldsymbol{\beta}$ and $\sigma$ is

$$(3.31) \qquad p(\boldsymbol{\beta}, \sigma|\mathbf{y}, X) \propto \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{1}{2\sigma^2}[\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]\right\}.$$



From (3.31) it is seen immediately that the conditional posterior pdf for $\boldsymbol{\beta}$, given $\sigma$ [i.e., $p(\boldsymbol{\beta}|\sigma, \mathbf{y}, X)$], is a $k$-dimensional multivariate normal pdf with mean $\hat{\boldsymbol{\beta}}$ and covariance matrix $(X'X)^{-1}\sigma^2$. Although this fact is interesting and useful in certain derivations, $\sigma^2$ is rarely known in practice and thus the conditional covariance matrix $(X'X)^{-1}\sigma^2$ cannot be evaluated. To get rid of the troublesome parameter $\sigma$, we integrate (3.31) with respect to $\sigma$ to obtain the following marginal posterior pdf for the elements of $\boldsymbol{\beta}$:

$$(3.32) \qquad \begin{aligned} p(\boldsymbol{\beta}|\mathbf{y}, X) &= \int_0^\infty p(\boldsymbol{\beta}, \sigma|\mathbf{y}, X)\, d\sigma \\ &\propto \{\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\}^{-n/2}, \end{aligned}$$

which is in the form of a multivariate Student $t$ pdf. This posterior pdf serves as a basis for making inferences about $\boldsymbol{\beta}$. Before turning to further analysis of it, we note that the marginal posterior pdf for $\sigma$ can be obtained from (3.31) by integrating with respect to the elements of $\boldsymbol{\beta}$; that is,

$$(3.33) \qquad \begin{aligned} p(\sigma|\mathbf{y}, X) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\boldsymbol{\beta}, \sigma|\mathbf{y}, X)\, d\boldsymbol{\beta} \\ &\propto \frac{1}{\sigma^{\nu+1}} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right), \end{aligned}$$

which is in the form of an inverted gamma pdf and exactly in the same form as (3.15), except that here $\nu = n - k$. By simple changes of variable the posterior pdf's for $\sigma^2$ or $h = 1/\sigma^2$ can be obtained from (3.33) if they are wanted.

We now return to the analysis of (3.32), the marginal posterior pdf for $\boldsymbol{\beta}$. First, we derive the marginal posterior pdf for a single element of $\boldsymbol{\beta}$, say $\beta_1$. This can be done in two ways, namely, by integrating (3.31) with respect to $\beta_2, \beta_3, \ldots, \beta_k$ and then with respect to $\sigma$ or by integrating (3.32) with respect to $\beta_2, \beta_3, \ldots, \beta_k$. We take the second route here. For convenience rewrite (3.32) as follows:

$$(3.34) \qquad p(\boldsymbol{\delta}|\mathbf{y}, X) \propto (\nu + \boldsymbol{\delta}'H\boldsymbol{\delta})^{-n/2},$$

with $\boldsymbol{\delta}' = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'$ and $H = X'X/s^2$. Now let $\delta_1 = \beta_1 - \hat{\beta}_1$, a scalar, and $\boldsymbol{\delta}_2' = (\beta_2 - \hat{\beta}_2, \beta_3 - \hat{\beta}_3, \ldots, \beta_k - \hat{\beta}_k)$. Then

$$(3.35) \qquad \boldsymbol{\delta}'H\boldsymbol{\delta} = \delta_1^2 h_{11} + \boldsymbol{\delta}_2'H_{22}\boldsymbol{\delta}_2 + 2\delta_1 H_{12}\boldsymbol{\delta}_2,$$

where $H$ has been partitioned to conform with the partitioning of $\boldsymbol{\delta}$; that is

$$H = \left(\begin{array}{c|c} h_{11} & H_{12} \\ \hline H_{21} & H_{22} \end{array}\right),$$

with $h_{11}$, a scalar, $H_{12}$ a $1 \times (k - 1)$ vector, $H_{21}$ a $(k - 1) \times 1$ vector and $H_{22}$ a $(k - 1) \times (k - 1)$ matrix. Now complete the square on $\boldsymbol{\delta}_2$ in (3.35) as follows

$$(3.36)\quad \begin{aligned}\boldsymbol{\delta}'H\boldsymbol{\delta} &= \delta_1^2h_{11} - \delta_1^2H_{12}H_{22}^{-1}H_{21} \\ &\quad + (\boldsymbol{\delta}_2 + \delta_1H_{22}^{-1}H_{21})'H_{22}(\boldsymbol{\delta}_2 + \delta_1H_{22}^{-1}H_{21}) \\ &= \delta_1^2(h_{11} - H_{12}H_{22}^{-1}H_{21}) \\ &\quad + (\boldsymbol{\delta}_2 + \delta_1H_{22}^{-1}H_{21})'H_{22}(\boldsymbol{\delta}_2 + \delta_1H_{22}^{-1}H_{21}).\end{aligned}$$

Then substitute from (3.36) in (3.34) to obtain

$$(3.37)\quad \begin{aligned}p(\delta_1, \boldsymbol{\delta}_2|\mathbf{y}, X) &\propto \left(\nu + \frac{\delta_1^2}{h^{11}}\right)^{-n/2} \\ &\quad \times [1 + (\boldsymbol{\delta}_2 + \delta_1H_{22}^{-1}H_{21})'C(\boldsymbol{\delta}_2 + \delta_1H_{22}^{-1}H_{21})]^{-n/2},\end{aligned}$$

with $C = H_{22}/(\nu + \delta_1^2/h^{11})$, where $h^{11} = (h_{11} - H_{12}H_{22}^{-1}H_{21})^{-1}$, the (1, 1) element of the inverse of $H$. Now (3.37) can be integrated with respect to $\boldsymbol{\delta}_2$ by using the properties of the multivariate Student $t$ pdf to yield[9]:

$$(3.38)\quad \begin{aligned}p(\delta_1|\mathbf{y}, X) &\propto \left(\nu + \frac{\delta_1^2}{h^{11}}\right)^{-n/2}|C|^{-1/2} \\ &\propto \left(\nu + \frac{\delta_1^2}{h^{11}}\right)^{-(n-k+1)/2}.\end{aligned}$$

Thus

$$(3.39)\quad \frac{\delta_1}{(h^{11})^{1/2}} = \frac{\beta_1 - \hat{\beta}_1}{s(m^{11})^{1/2}} \sim t_\nu,$$

where $m^{11}$ is the (1, 1) element of $(X'X)^{-1}$. Since the choice of which element of $\boldsymbol{\beta}$ is labeled $\beta_1$ is open, we have for the $i$th element of $\boldsymbol{\beta}$:

$$(3.40)\quad \frac{\delta_i}{(h^{ii})^{1/2}} = \frac{\beta_i - \hat{\beta}_i}{s(m^{ii})^{1/2}} \sim t_\nu.$$

This result enables us to make inferences about $\beta_i$ conveniently by consulting the $t$ tables for $\nu = n - k$ degrees of freedom. Note also that a simple change of variable in (3.38), namely, $F = \delta_1^2/h^{11} = (\beta_1 - \hat{\beta}_1)^2/s^2m^{11}$, yields

$$(3.41)\quad P(F) \propto F^{-1/2}(\nu + F)^{-(\nu+1)/2}, \qquad 0 < F < \infty,$$

and thus $F = (\beta_1 - \hat{\beta}_1)^2/s^2m^{11}$ has the $F_{1,\nu}$ pdf.

If we are interested in the marginal posterior pdf of a subset of elements of $\boldsymbol{\beta}$ (or of $\boldsymbol{\delta} = \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}$), we can partition $\boldsymbol{\delta}' = (\boldsymbol{\delta}_1' \vdots \boldsymbol{\delta}_2')$ and write the quadratic form $\boldsymbol{\delta}'H\boldsymbol{\delta}$ appearing in (3.34) as follows:

$$(3.42)\quad \begin{aligned}\boldsymbol{\delta}'H\boldsymbol{\delta} &= (\boldsymbol{\delta}_1' \vdots \boldsymbol{\delta}_2')\begin{pmatrix}H_{11} & H_{12} \\ H_{21} & H_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2\end{pmatrix} \\ &= \boldsymbol{\delta}_1'H_{11}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2'H_{22}\boldsymbol{\delta}_2 + 2\boldsymbol{\delta}_1'H_{12}\boldsymbol{\delta}_2,\end{aligned}$$

[9] The integration of $|C|^{1/2}[1 + (\boldsymbol{\delta}_2 + \delta_1H_{22}^{-1}H_{21})'C(\boldsymbol{\delta}_2 + \delta_1H_{22}^{-1}H_{21})]^{-n/2}$ with respect to $\boldsymbol{\delta}_2$ yields a constant independent of $\delta_1$. Since $C$ involves $\delta_1$, it must appear in (3.38).



where the partitioning of $H$ has been done to conform with that of $\boldsymbol{\delta}$. On completing the square on $\boldsymbol{\delta}_2$ in (3.42) we have

$$\begin{aligned}\boldsymbol{\delta}'H\boldsymbol{\delta} &= \boldsymbol{\delta}_1'H_{11}\boldsymbol{\delta}_1 - \boldsymbol{\delta}_1'H_{12}H_{22}^{-1}H_{21}\boldsymbol{\delta}_1 \\ &\quad + (\boldsymbol{\delta}_2 + H_{22}^{-1}H_{21}\boldsymbol{\delta}_1)'H_{22}(\boldsymbol{\delta}_2 + H_{22}^{-1}H_{21}\boldsymbol{\delta}_1),\end{aligned}$$

which, when substituted in (3.34), yields

$$(3.43)\quad \begin{aligned}p(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2|\mathbf{y}, X) &\propto [\nu + \boldsymbol{\delta}_1'(H_{11} - H_{12}H_{22}^{-1}H_{21})\boldsymbol{\delta}_1 \\ &\quad + (\boldsymbol{\delta}_2 + H_{22}^{-1}H_{21}\boldsymbol{\delta}_1)'H_{22}(\boldsymbol{\delta}_2 + H_{22}^{-1}H_{21}\boldsymbol{\delta}_1)]^{-n/2}.\end{aligned}$$

As an aside, we see from (3.43) that the conditional pdf for $\boldsymbol{\delta}_2$, given $\boldsymbol{\delta}_1$, $p(\boldsymbol{\delta}_2|\boldsymbol{\delta}_1, \mathbf{y}, X)$, is in the multivariate Student $t$ form with conditional mean $-H_{22}^{-1}H_{21}\boldsymbol{\delta}_1$. Thus, since $\boldsymbol{\delta}_2 = \boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2$, the conditional mean of $\boldsymbol{\beta}_2$, given $\boldsymbol{\beta}_1$, is $\hat{\boldsymbol{\beta}}_2 - H_{22}^{-1}H_{21}(\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1)$.

To obtain the marginal posterior pdf for $\boldsymbol{\delta}_1$, we have to integrate (3.43) with respect to $\boldsymbol{\delta}_2$ which can be done by using properties of the multivariate Student $t$ pdf. This operation results in[10]

$$(3.44)\quad p(\boldsymbol{\delta}_1|\mathbf{y}, X) \propto [\nu + \boldsymbol{\delta}_1'(H_{11} - H_{12}H_{22}^{-1}H_{21})\boldsymbol{\delta}_1]^{-(n-k_2)/2},$$

where $k_2$ is the number of elements in $\boldsymbol{\delta}_2$. Note that $n - k_2 = n - k + k - k_2 = \nu + k_1$ so that the exponent of (3.44) is $-(\nu + k_1)/2$. Thus from (3.44) the marginal posterior pdf for $\boldsymbol{\delta}_1$ is in the multivariate Student $t$ form with $E\boldsymbol{\delta}_1 = E(\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1) = \mathbf{0}$, that is $E\boldsymbol{\beta}_1 = \hat{\boldsymbol{\beta}}_1$, $\nu > 1$, and $E\boldsymbol{\delta}_1\boldsymbol{\delta}_1' = E(\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1)(\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1)' = [(\nu/(\nu - 2)](H_{11} - H_{12}H_{22}^{-1}H_{21})^{-1}$, $\nu > 2$, where it is to be remembered that the matrices $H_{\alpha l}$, $\alpha, l = 1, 2$, are submatrices of $H = X'X/s^2$.

Next, using the prior assumptions in (3.30), we turn to the problem of deriving the posterior pdf of a linear combination of the elements of $\boldsymbol{\beta}$, say $\alpha = \mathbf{l}'\boldsymbol{\beta}$, where $\alpha$ is a scalar parameter and $\mathbf{l}'$ is a $1 \times k$ vector of fixed numbers; for example, if the elements of $\boldsymbol{\beta}$ are Cobb-Douglas production function parameters, we might be interested in $\alpha = \beta_1 + \beta_2 + \cdots + \beta_k$, which is the "returns to scale" parameter. In this application $\mathbf{l}' = (1, 1, \ldots, 1)$. In other situations other linear combinations of the elements of $\boldsymbol{\beta}$ may be of interest. To obtain the posterior pdf of $\alpha = \mathbf{l}'\boldsymbol{\beta}$ we note that the joint posterior pdf for $\boldsymbol{\beta}$ and $\sigma$ can be written as

$$(3.45)\quad p(\boldsymbol{\beta}, \sigma|\mathbf{y}, X) = p(\boldsymbol{\beta}|\sigma, \mathbf{y}, X)\,p(\sigma|\mathbf{y}, X),$$

[10] Write (3.43) as

$$\begin{aligned}p(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2|\mathbf{y}, X) &\propto \{[\nu + \boldsymbol{\delta}_1'(H_{11} - H_{12}H_{22}^{-1}H_{21})\boldsymbol{\delta}_1]^{-n/2}|C|^{-1/2}\} \\ &\quad \times \{|C|^{1/2}[1 + (\boldsymbol{\delta}_2 + H_{22}^{-1}H_{21}\boldsymbol{\delta}_1)'C(\boldsymbol{\delta}_2 + H_{22}^{-1}H_{21}\boldsymbol{\delta}_1)]^{-n/2}\},\end{aligned}$$

with $C \equiv H_{22}/[\nu + \boldsymbol{\delta}_1'(H_{11} - H_{12}H_{22}^{-1}H_{21})\boldsymbol{\delta}_1]$. By integrating over $\boldsymbol{\delta}_2$ the second factor yields a constant independent of $\boldsymbol{\delta}_1$ and the first factor is proportional to the expression shown in (3.44).

with $p(\boldsymbol{\beta}|\sigma, \mathbf{y}, X)$ in the multivariate normal form. Thus, conditional on $\sigma$, $\alpha = \mathbf{l}'\boldsymbol{\beta}$ will be normally distributed, since it is a linear combination of normally distributed variables, the elements of $\boldsymbol{\beta}$, with mean $\hat{\alpha} = \mathbf{l}'\hat{\boldsymbol{\beta}}$, and conditional posterior variance

$$E[(\alpha - \hat{\alpha})^2|\sigma, \mathbf{y}, X)] = E[\mathbf{l}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{l}|\sigma, \mathbf{y}, X] = \mathbf{l}'(X'X)^{-1}\mathbf{l}\sigma^2,$$

since the conditional covariance matrix for $\boldsymbol{\beta}$, given $\sigma$, is $(X'X)^{-1}\sigma^2$. Thus the marginal posterior pdf for $\alpha$ can be obtained by integrating $p(\alpha, \sigma|\mathbf{y}, X) = p(\alpha|\sigma, \mathbf{y}, X)p(\sigma|\mathbf{y}, X)$ with respect to $\sigma$. Letting $c = \mathbf{l}'(X'X)^{-1}\mathbf{l}$, we have

$$p(\alpha|\sigma, \mathbf{y}, X) \propto \frac{1}{\sigma} \exp\left[-\frac{(\alpha - \hat{\alpha})^2}{2\sigma^2 c}\right],$$

$$p(\sigma|\mathbf{y}, X) \propto \frac{1}{\sigma^{\nu+1}} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right),$$

and

$$\begin{aligned} p(\alpha|\mathbf{y}, X) &= \int_0^\infty p(\alpha|\sigma, \mathbf{y}, X)\, p(\sigma|\mathbf{y}, X)\, d\sigma \\ &\propto \int_0^\infty \frac{1}{\sigma^{\nu+2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + \frac{(\alpha - \hat{\alpha})^2}{c}\right]\right\} d\sigma \qquad (3.46) \\ &\propto \left[\nu + \frac{(\alpha - \hat{\alpha})^2}{s^2 c}\right]^{-(\nu+1)/2}. \end{aligned}$$

Thus $\alpha$ has a posterior pdf in the univariate Student $t$ form; that is,

$$\frac{\alpha - \hat{\alpha}}{sc^{1/2}} \sim t_\nu, \qquad (3.47)$$

with $\nu = n - k$, $\hat{\alpha} = \mathbf{l}'\hat{\boldsymbol{\beta}}$ and $c = \mathbf{l}'(X'X)^{-1}\mathbf{l}$. This fact can be utilized to make inferences about $\alpha$.[11]

### 3.2.3 Posterior Pdf Based on an Informative Prior Pdf

We next take up the problem of using the posterior pdf in (3.31) as a prior pdf in the analysis of a new sample of data generated by the same regression process. To distinguish between the two samples subscripts 1 and 2 are employed. With this notation the posterior pdf in (3.31), which we use as a prior pdf for analysis of a new sample, is

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma|\mathbf{y}_1, X_1) &\propto \frac{1}{\sigma^{n_1+1}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}_1 - X_1\boldsymbol{\beta})'(\mathbf{y}_1 - X_1\boldsymbol{\beta})\right] \qquad (3.48) \\ &\propto \frac{1}{\sigma^{n_1+1}} \exp\left\{-\frac{1}{2\sigma^2}[\nu_1 s_1^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_1)'X_1'X_1(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_1)]\right\}, \end{aligned}$$

[11] See, Appendix B for the derivation of the joint distribution of several linear combinations of variables having a multivariate Student pdf.



with $\nu_1 = n_1 - k$, $\hat{\boldsymbol{\beta}}_1 = (X_1'X_1)^{-1}X_1'\mathbf{y}_1$ and $\nu_1 s_1^2 = (\mathbf{y}_1 - X_1\hat{\boldsymbol{\beta}}_1)'(\mathbf{y}_1 - X_1\hat{\boldsymbol{\beta}}_1)$. Viewing (3.48) as a prior pdf, we see that it factors into a normal part for $\boldsymbol{\beta}$, given $\sigma$, with mean $\hat{\boldsymbol{\beta}}_1$ and covariance matrix $(X_1'X_1)^{-1}\sigma^2$ and a marginal pdf for $\sigma$ in the inverted gamma form with parameters $\nu_1$ and $s_1^2$; that is, from the second line of (3.48)

$$p(\boldsymbol{\beta}|\sigma, \hat{\boldsymbol{\beta}}_1, s_1^2) \propto \frac{1}{\sigma^k} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_1)'X_1'X_1(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_1)\right]$$

and

$$p(\sigma|\hat{\boldsymbol{\beta}}_1, s_1^2) \propto \frac{1}{\sigma^{\nu_1+1}} \exp\left(-\frac{\nu_1 s_1^2}{2\sigma^2}\right).$$

Thus the prior pdf's parameters are just the quantities $\hat{\boldsymbol{\beta}}_1$, $s_1^2$, $X_1'X_1$, and $\nu_1$.

The likelihood function for the second sample $(\mathbf{y}_2, X_2)$, where $\mathbf{y}_2$ is an $n_2 \times 1$ vector of observations on the dependent variable in the second sample and $X_2$ is an $n_2 \times k$ matrix with rank $k$ of observations on the $k$ independent variables in the second sample, is assumed to be given by

$$l(\boldsymbol{\beta}, \sigma|\mathbf{y}_2, X_2) \propto \frac{1}{\sigma^{n_2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}_2 - X_2\boldsymbol{\beta})'(\mathbf{y}_2 - X_2\boldsymbol{\beta})\right]. \qquad (3.49)$$

Note that $\boldsymbol{\beta}$ and $\sigma$ for the second sample are assumed to be the same parameters as those for the first sample.

On combining the prior pdf in (3.48) with the likelihood function in (3.49), we obtain the posterior pdf:

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma|\mathbf{y}_1, \mathbf{y}_2, X_1, X_2) \propto{}& \frac{1}{\sigma^{n_1+n_2+1}} \\ &\times \exp\left\{-\frac{1}{2\sigma^2}[(\mathbf{y}_1 - X_1\boldsymbol{\beta})'(\mathbf{y}_1 - X_1\boldsymbol{\beta}) \right. \qquad (3.50) \\ &\left. + (\mathbf{y}_2 - X_2\boldsymbol{\beta})'(\mathbf{y}_2 - X_2\boldsymbol{\beta})]\right\}. \end{aligned}$$

This expression can be brought into a more convenient form on completing the square in the exponent; that is,

$$(\mathbf{y}_1 - X_1\boldsymbol{\beta})'(\mathbf{y}_1 - X_1\boldsymbol{\beta}) + (\mathbf{y}_2 - X_2\boldsymbol{\beta})'(\mathbf{y}_2 - X_2\boldsymbol{\beta}) = \nu s^2 + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'M(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}),$$

where

$$\begin{aligned} M &= X_1'X_1 + X_2'X_2, \\ \tilde{\boldsymbol{\beta}} &= M^{-1}(X_1'\mathbf{y}_1 + X_2'\mathbf{y}_2), \\ \nu s^2 &= (\mathbf{y}_1 - X_1\tilde{\boldsymbol{\beta}})'(\mathbf{y}_1 - X_1'\tilde{\boldsymbol{\beta}}) + (\mathbf{y}_2 - X_2\tilde{\boldsymbol{\beta}})'(\mathbf{y}_2 - X_2\tilde{\boldsymbol{\beta}}), \\ \nu &= n_1 + n_2 - k. \end{aligned}$$

Thus (3.50) can be written as

$$p(\boldsymbol{\beta}, \sigma|\mathbf{y}_1, \mathbf{y}_2, X_1, X_2) \propto \frac{1}{\sigma^{n_1+n_2+1}} \exp\left\{-\frac{1}{2\sigma^2}[\nu s^2 + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'M(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})]\right\}. \qquad (3.51)$$

It is seen that (3.51) is in exactly the same form as (3.31) and thus it can be analyzed by using exactly the same techniques. Further, if we had pooled the two samples, based our likelihood function on both samples, and used the diffuse prior pdf in (3.30), the resulting posterior pdf would be exactly (3.50), from which (3.51) can be derived.