Appendix C

The Matrix Approach to Linear Regression Model

This appendix presents the classical linear regression model involving k variables (Y and $X_2, X_3, \ldots, X_k$) in matrix algebra notation. Conceptually, the k-variable model is a logical extension of the two- and three-variable models considered thus far in this text. Therefore, this appendix presents very few new concepts save for the matrix notation.[1]

A great advantage of matrix algebra over scalar algebra (elementary algebra dealing with scalars or real numbers) is that it provides a compact method of handling regression models involving any number of variables; once the k-variable model is formulated and solved in matrix notation, the solution applies to one, two, three, or any number of variables.

C.1 The k-Variable Linear Regression Model

If we generalize the two- and three-variable linear regression models, the k-variable population regression function (PRF) model involving the dependent variable Y and $k-1$ explanatory variables $X_2, X_3, \ldots, X_k$ may be written as

$$\text{PRF:}\quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \qquad i = 1, 2, 3, \ldots, n \tag{C.1.1}$$

where $\beta_1 =$ the intercept, β_2 to $\beta_k =$ partial slope coefficients, $u =$ stochastic disturbance term, and $i = i$th observation, n being the size of the population. The PRF (C.1.1) is to be interpreted in the usual manner: It gives the mean or expected value of Y conditional upon the fixed (in repeated sampling) values of $X_2, X_3, \ldots, X_k$, that is, $E(Y \mid X_{2i}, X_{3i}, \ldots, X_{ki})$.

[1]Readers not familiar with matrix algebra should review **Appendix B** before proceeding any further. **Appendix B** provides the essentials of matrix algebra needed to follow this appendix.

Equation (C.1.1) is a shorthand expression for the following set of n simultaneous equations:

$$
\begin{aligned}
Y_1 &= \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \cdots + \beta_k X_{k1} + u_1 \\
Y_2 &= \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \cdots + \beta_k X_{k2} + u_2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_n &= \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \cdots + \beta_k X_{kn} + u_n
\end{aligned}
\tag{C.1.2}
$$

Let us write the system of equations (C.1.2) in an alternative but more illuminating way as follows:[2]

$$
\begin{array}{ccccccc}
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} & = & \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix} & \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} & + & \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \\
\mathbf{y} & = & \mathbf{X} & \boldsymbol{\beta} & + & \mathbf{u} \\
n \times 1 & & n \times k & k \times 1 & & n \times 1
\end{array}
\tag{C.1.3}
$$

where $\mathbf{y} = n \times 1$ column vector of observations on the dependent variable Y
$\mathbf{X} = n \times k$ matrix giving n observations on $k - 1$ variables X_2 to X_k, the first column of 1's representing the intercept term (this matrix is also known as the **data matrix**)
$\boldsymbol{\beta} = k \times 1$ column vector of the unknown parameters $\beta_1, \beta_2, \ldots, \beta_k$
$\mathbf{u} = n \times 1$ column vector of n disturbances u_i

Using the rules of matrix multiplication and addition, the reader should verify that systems (C.1.2) and (C.1.3) are equivalent.

System (C.1.3) is known as the *matrix representation of the general (k-variable) linear regression model*. It can be written more compactly as

$$
\underset{n \times 1}{\mathbf{y}} = \underset{n \times k}{\mathbf{X}}\ \underset{k \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\mathbf{u}}
\tag{C.1.4}
$$

Where there is no confusion about the dimensions or orders of the matrix $\mathbf{X}$ and the vectors $\mathbf{y}$, $\boldsymbol{\beta}$, and $\mathbf{u}$, Eq. (C.1.4) may be written simply as

$$
\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}
\tag{C.1.5}
$$

As an illustration of the matrix representation, consider the two-variable consumption–income model considered in Chapter 3, namely, $Y_i = \beta_1 + \beta_2 X_i + u_i$, where Y is consumption expenditure and X is income. Using the data given in Table 3.2, we may write the

[2]Following the notation introduced in **Appendix B,** we shall represent vectors by lowercase boldfaced letters and matrices by uppercase boldfaced letters.

matrix formulation as

$$\begin{bmatrix} 70 \\ 65 \\ 90 \\ 95 \\ 110 \\ 115 \\ 120 \\ 140 \\ 155 \\ 150 \end{bmatrix} = \begin{bmatrix} 1 & 80 \\ 1 & 100 \\ 1 & 120 \\ 1 & 140 \\ 1 & 160 \\ 1 & 180 \\ 1 & 200 \\ 1 & 220 \\ 1 & 240 \\ 1 & 260 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \end{bmatrix} \tag{C.1.6}$$

$$\underset{10 \times 1}{\mathbf{y}} = \underset{10 \times 2}{\mathbf{X}} \quad \underset{2 \times 1}{\boldsymbol{\beta}} + \underset{10 \times 1}{\mathbf{u}}$$

As in the two- and three-variable cases, our objective is to estimate the parameters of the multiple regression (C.1.1) and to draw inferences about them from the data at hand. In matrix notation this amounts to estimating $\boldsymbol{\beta}$ and drawing inferences about this $\boldsymbol{\beta}$. For the purpose of estimation, we may use the method of ordinary least squares (OLS) or the method of maximum likelihood (ML). But as noted before, these two methods yield identical estimates of the regression coefficients.[3] Therefore, we shall confine our attention to the method of OLS.

C.2 Assumptions of the Classical Linear Regression Model in Matrix Notation

The assumptions underlying the classical linear regression model are given in Table C.1; they are presented both in scalar notation and in matrix notation. Assumption 1 given in Eq. (C.2.1) means that the expected value of the disturbance vector $\mathbf{u}$, that is, of each of its elements, is zero. More explicitly, $E(\mathbf{u}) = \mathbf{0}$ means

$$E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{C.2.1}$$

Assumption 2 (Eq. [C.2.2]) is a compact way of expressing the two assumptions given in Eqs. (3.2.5) and (3.2.2) by the scalar notation. To see this, we can write

$$E(\mathbf{u}\mathbf{u}') = E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}[u_1 \quad u_2 \quad \cdots \quad u_n]$$

[3]The proof that this is so in the *k*-variable case can be found in the footnote reference given in Chapter 4.

TABLE C.1
Assumptions of the Classical Linear Regression Model

Scalar Notation		Matrix Notation
1. $E(u_i) = 0$, for each i	(3.2.1)	1. $E(\mathbf{u}) = \mathbf{0}$ where $\mathbf{u}$ and $\mathbf{0}$ are $n \times 1$ column vectors, $\mathbf{0}$ being a null vector
2. $E(u_i u_j) = 0 \quad i \neq j$ $= \sigma^2 \quad i = j$	(3.2.5) (3.2.2)	2. $E(\mathbf{uu}') = \sigma^2\mathbf{I}$ where $\mathbf{I}$ is an $n \times n$ identity matrix
3. $X_2, X_3, \ldots, X_k$ are nonstochastic or fixed		3. The $n \times k$ matrix $\mathbf{X}$ is nonstochastic, that is, it consists of a set of fixed numbers
4. There is no exact linear relationship among the X variables, that is, no multicollinearity	(7.1.9)	4. The rank of $\mathbf{X}$ is $p(\mathbf{X}) = k$, where k is the number of columns in $\mathbf{X}$ and k is less than the number of observations, n
5. For hypothesis testing, $u_i \sim N(0, \sigma^2)$	(4.2.4)	5. The $\mathbf{u}$ vector has a multivariate normal distribution, i.e., $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$

where $\mathbf{u}'$ is the transpose of the column vector $\mathbf{u}$, or a row vector. Performing the multiplication, we obtain

$$E(\mathbf{uu}') = E\begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \cdots & \cdots & \cdots & \cdots \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix}$$

Applying the expectations operator E to each element of the preceding matrix, we obtain

$$E(\mathbf{uu}') = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \cdots & \cdots & \cdots & \cdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{bmatrix} \tag{C.2.2}$$

Because of the assumptions of homoscedasticity and no serial correlation, matrix (C.2.2) reduces to

$$E(\mathbf{uu}') = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma^2\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{C.2.3}$$

$$= \sigma^2\mathbf{I}$$

where $\mathbf{I}$ is an $n \times n$ identity matrix.

Matrix (C.2.2) (and its representation given in Eq. [C.2.3]) is called the **variance-covariance matrix** of the disturbances u_i; the elements on the main diagonal of this matrix (running from the upper left corner to the lower right corner) give the variances, and the elements off the main diagonal give the covariances.[4] Note that the variance–covariance matrix is **symmetric:** The elements above and below the main diagonal are reflections of one another.

Assumption 3 in Table C.1 states that the $n \times k$ matrix $\mathbf{X}$ is nonstochastic; that is, it consists of fixed numbers. As noted previously, our regression analysis is conditional regression analysis, conditional upon the fixed values of the X variables.

Assumption 4 states that the $\mathbf{X}$ matrix has full column rank equal to k, the number of columns in the matrix. This means that the columns of the X matrix are linearly independent; that is, there is no **exact linear relationship** among the X variables. In other words there is no multicollinearity. In scalar notation this is equivalent to saying that there exists no set of numbers $\lambda_1, \lambda_2, \ldots, \lambda_k$ not all zero such that (cf. Eq. [7.1.8])

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} + \cdots + \lambda_k X_{ki} = 0 \tag{C.2.4}$$

where $X_{1i} = 1$ for all i (to allow for the column of 1's in the $\mathbf{X}$ matrix). In matrix notation, Eq. (C.2.4) can be represented as

$$\boldsymbol{\lambda}'\mathbf{x} = 0 \tag{C.2.5}$$

where $\boldsymbol{\lambda}'$ is a $1 \times k$ row vector and $\mathbf{x}$ is a $k \times 1$ column vector.

If an exact linear relationship such as Eq. (C.2.4) exists, the variables are said to be collinear. If, on the other hand, Eq. (C.2.4) holds true only if $\lambda_1 = \lambda_2 = \lambda_3 = \cdots = 0$, then the X variables are said to be linearly independent. An intuitive reason for the *no multicollinearity* assumption was given in Chapter 7, and we explored this assumption further in Chapter 10.

C.3 OLS Estimation

To obtain the OLS estimate of $\boldsymbol{\beta}$, let us first write the k-variable sample regression function (SRF):

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i \tag{C.3.1}$$

which can be written more compactly in matrix notation as

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}} \tag{C.3.2}$$

and in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} \tag{C.3.3}$$

$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times k}{\mathbf{X}} \quad \underset{k \times 1}{\hat{\boldsymbol{\beta}}} + \underset{n \times 1}{\hat{\mathbf{u}}}$$

where $\hat{\boldsymbol{\beta}}$ is a k-element column vector of the OLS estimators of the regression coefficients and where $\hat{\mathbf{u}}$ is an $n \times 1$ column vector of n residuals.

[4] By definition, the variance of $u_i = E[u_i - E(u_i)]^2$ and the covariance between u_i and $u_j = E[u_i - E(u_i)][u_j - E(u_j)]$. But because of the assumption $E(u_i) = 0$ for each i, we have the variance-covariance matrix (C.2.3).

As in the two- and three-variable models, in the k-variable case the OLS estimators are obtained by minimizing

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2 \tag{C.3.4}$$

where $\sum \hat{u}_i^2$ is the residual sum of squares (RSS). In matrix notation, this amounts to minimizing $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ since

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = [\hat{u}_1 \quad \hat{u}_2 \quad \cdots \quad \hat{u}_n] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_n^2 = \sum \hat{u}_i^2 \tag{C.3.5}$$

Now from Eq. (C.3.2) we obtain

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \tag{C.3.6}$$

Therefore,

$$\begin{aligned} \hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned} \tag{C.3.7}$$

where use is made of the properties of the transpose of a matrix, namely, $(\mathbf{X}\hat{\boldsymbol{\beta}})' = \hat{\boldsymbol{\beta}}'\mathbf{X}'$; and since $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ is a scalar (a real number), it is equal to its transpose $\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$.

Equation (C.3.7) is the matrix representation of (C.3.4). In scalar notation, the method of OLS consists in so estimating $\beta_1, \beta_2, \ldots, \beta_k$ that $\sum \hat{u}_i^2$ is as small as possible. This is done by differentiating Eq. (C.3.4) partially with respect to $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ and setting the resulting expressions to zero. This process yields k simultaneous equations in k unknowns, the normal equations of the least-squares theory. As shown in Appendix CA, Section CA.1, these equations are as follows:

$$\begin{aligned} n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki} &= \sum Y_i \\ \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i}X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i}X_{ki} &= \sum X_{2i}Y_i \\ \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{3i}X_{2i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{3i}X_{ki} &= \sum X_{3i}Y_i \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{ki}X_{2i} + \hat{\beta}_3 \sum X_{ki}X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki}^2 &= \sum X_{ki}Y_i \end{aligned} \tag{C.3.8}$$

(C.3.8)[5]

In matrix form, Eq. (C.3.8) can be represented as

$$\underbrace{\begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \cdots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \cdots & \sum X_{2i}X_{ki} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 & \cdots & \sum X_{3i}X_{ki} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \cdots & \sum X_{ki}^2 \end{bmatrix}}_{(\mathbf{X}'\mathbf{X})} \underbrace{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}}_{\hat{\boldsymbol{\beta}}} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & \cdots & X_{2n} \\ X_{31} & X_{32} & \cdots & X_{3n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix}}_{\mathbf{X}'} \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{y}} \tag{C.3.9}$$

[5]These equations can be remembered easily. Start with the equation $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki}$. Summing this equation over the n values gives the first equation in (C.3.8); multiplying it by X_2 on both sides and summing over n gives the second equation; multiplying it by X_3 on both sides and summing over n gives the third equation; and so on. In passing, note that the first equation in (C.3.8) gives at once $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \cdots - \hat{\beta}_k \bar{X}_k$ (cf. [7.4.6]).

or, more compactly, as

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{C.3.10}$$

Note these features of the $(\mathbf{X}'\mathbf{X})$ matrix: (1) It gives the raw sums of squares and cross products of the X variables, one of which is the intercept term taking the value of 1 for each observation. The elements on the main diagonal give the raw sums of squares, and those off the main diagonal give the raw sums of cross products (by *raw* we mean in original units of measurement). (2) It is symmetrical since the cross product between X_{2i} and X_{3i} is the same as that between X_{3i} and X_{2i}. (3) It is of order $(k \times k)$, that is, k rows and k columns.

In Eq. (C.3.10) the known quantities are $(\mathbf{X}'\mathbf{X})$ and $(\mathbf{X}'\mathbf{y})$ (the cross product between the X variables and y) and the unknown is $\hat{\boldsymbol{\beta}}$. Now using matrix algebra, if the inverse of $(\mathbf{X}'\mathbf{X})$ exists, say, $(\mathbf{X}'\mathbf{X})^{-1}$, then premultiplying both sides of Eq. (C.3.10) by this inverse, we obtain

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

But since $(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{I}$, an identity matrix of order $k \times k$, we get

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

or

$$\underset{k \times 1}{\hat{\boldsymbol{\beta}}} = \underset{k \times k}{(\mathbf{X}'\mathbf{X})^{-1}} \; \underset{(k \times n)}{\mathbf{X}'} \; \underset{(n \times 1)}{\mathbf{y}} \tag{C.3.11}$$

Equation (C.3.11) is a fundamental result of the OLS theory in matrix notation. It shows how the $\hat{\boldsymbol{\beta}}$ vector can be estimated from the given data. Although Eq. (C.3.11) was obtained from Eq. (C.3.9), it can be obtained directly from Eq. (C.3.7) by differentiating $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ with respect to $\hat{\boldsymbol{\beta}}$. The proof is given in Appendix CA, Section CA.2.

An Illustration

As an illustration of the matrix methods developed so far, let us work a consumption–income example using the data in Eq. (C.1.6). For the two-variable case we have

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ & \cdots \\ 1 & X_N \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

and

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & X_3 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Using the data given in Eq. (C.1.6), we obtain

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 10 & 1700 \\ 1700 & 322000 \end{bmatrix}$$

and

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1110 \\ 205500 \end{bmatrix}$$

Using the rules of matrix inversion given in **Appendix B, Section B.3,** we can see that the inverse of the preceding $(\mathbf{X}'\mathbf{X})$ matrix is

$$\mathbf{X}'\mathbf{X}^{-1} = \begin{bmatrix} 0.97576 & -0.005152 \\ -0.005152 & 0.0000303 \end{bmatrix}$$

Therefore,

$$\begin{aligned}\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} &= \begin{bmatrix} 0.97576 & -0.005152 \\ -0.005152 & 0.0000303 \end{bmatrix} \begin{bmatrix} 1110 \\ 205500 \end{bmatrix} \\ &= \begin{bmatrix} 24.4545 \\ 0.5079 \end{bmatrix}\end{aligned}$$

Using the computer, we obtained $\hat{\beta}_1 = 24.4545$ and $\hat{\beta}_2 = 0.5091$. The difference between the two estimates is due to the rounding errors. In passing, note that in working on a desk calculator it is essential to obtain results to several significant digits to minimize the rounding errors.

Variance-Covariance Matrix of $\hat{\boldsymbol{\beta}}$

Matrix methods enable us to develop formulas not only for the variance of $\hat{\beta}_i$, any given element of $\hat{\boldsymbol{\beta}}$, but also for the covariance between any two elements of $\hat{\boldsymbol{\beta}}$, say, $\hat{\beta}_i$ and $\hat{\beta}_j$. We need these variances and covariances for the purpose of statistical inference.

By definition, the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is (compare Eq. [C.2.2])

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = E\{[\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})]'\}$$

which can be written explicitly as

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \text{var}\,(\hat{\beta}_1) & \text{cov}\,(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \text{cov}\,(\hat{\beta}_1, \hat{\beta}_k) \\ \text{cov}\,(\hat{\beta}_2, \hat{\beta}_1) & \text{var}\,(\hat{\beta}_2) & \cdots & \text{cov}\,(\hat{\beta}_2, \hat{\beta}_k) \\ \cdots & \cdots & \cdots & \cdots \\ \text{cov}\,(\hat{\beta}_k, \hat{\beta}_1) & \text{cov}\,(\hat{\beta}_k, \hat{\beta}_2) & \cdots & \text{var}\,(\hat{\beta}_k) \end{bmatrix} \tag{C.3.12}$$

It is shown in Appendix CA, Section CA.3, that the preceding variance-covariance matrix can be obtained from the following formula:

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{C.3.13}$$

where σ^2 is the homoscedastic variance of u_i and $(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse matrix appearing in Eq. (C.3.11), which gives the OLS estimator $\hat{\boldsymbol{\beta}}$.

In the two- and three-variable linear regression models an unbiased estimator of σ^2 was given by $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-2)$ and $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-3)$, respectively. In the k-variable case, the corresponding formula is

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum \hat{u}_i^2}{n-k} \\ &= \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}\end{aligned} \tag{C.3.14}$$

where there are now $n-k$ df. (Why?)

Although in principle $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ can be computed from the estimated residuals, in practice it can be obtained directly as follows. Recalling that $\sum \hat{u}_i^2$ $(= \text{RSS}) = \text{TSS} - \text{ESS}$, in the two-variable case we may write

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \tag{3.3.6}$$

and in the three-variable case

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{7.4.19}$$

By extending this principle, it can be seen that for the k-variable model

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \cdots - \hat{\beta}_k \sum y_i x_{ki} \tag{C.3.15}$$

In matrix notation,

$$\text{TSS: } \sum y_i^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2 \tag{C.3.16}$$

$$\text{ESS: } \hat{\beta}_2 \sum y_i x_{2i} + \cdots + \hat{\beta}_k \sum y_i x_{ki} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 \tag{C.3.17}$$

where the term $n\bar{Y}^2$ is known as the correction for mean.[6] Therefore,

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{C.3.18}$$

Once $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ is obtained, $\hat{\sigma}^2$ can be easily computed from Eq. (C.3.14), which, in turn, will enable us to estimate the variance-covariance matrix (C.3.13).

For our illustrative example,

$$\begin{aligned}\hat{\mathbf{u}}'\hat{\mathbf{u}} &= 132100 - [24.4545 \quad 0.5091]\begin{bmatrix} 1110 \\ 205500 \end{bmatrix} \\ &= 337.373\end{aligned}$$

Hence, $\hat{\sigma}^2 = (337.273/8) = 42.1591$, which is approximately the value obtained previously in Chapter 3.

[6]*Note:* $\sum y_i^2 = \sum(Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2$. Therefore, without the correction term, $\mathbf{y}'\mathbf{y}$ will give simply the raw sum of squares, not the sum of squared deviations.

Properties of OLS Vector $\hat{\boldsymbol{\beta}}$

In the two- and three-variable cases we know that the OLS estimators are linear and unbiased, and in the class of all linear unbiased estimators they have minimum variance (the Gauss–Markov property). In short, the OLS estimators are best linear unbiased estimators (BLUE). This property extends to the entire $\hat{\boldsymbol{\beta}}$ vector; that is, $\hat{\boldsymbol{\beta}}$ is linear (each of its elements is a linear function of Y, the dependent variable). $E(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}$, that is, the expected value of each element of $\hat{\boldsymbol{\beta}}$ is equal to the corresponding element of the true $\boldsymbol{\beta}$, and in the class of all linear unbiased estimators of $\boldsymbol{\beta}$, the OLS estimator $\hat{\boldsymbol{\beta}}$ has minimum variance.

The proof is given in Appendix CA, Section CA.4. As stated in the introduction, the k-variable case is in most cases a straight extension of the two- and three-variable cases.

C.4 The Coefficient of Determination R^2 in Matrix Notation

The coefficient of determination R^2 has been defined as

$$R^2 = \frac{\text{ESS}}{\text{TSS}}$$

In the two-variable case,

$$R^2 = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} \tag{3.5.6}$$

and in the three-variable case

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2} \tag{7.5.5}$$

Generalizing, we obtain for the k-variable case

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \cdots + \hat{\beta}_k \sum y_i x_{ki}}{\sum y_i^2} \tag{C.4.1}$$

By using Eqs. (C.3.16) and (C.3.17), Eq. (C.4.1) can be written as

$$R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} \tag{C.4.2}$$

which gives the matrix representation of R^2.

For our illustrative example,

$$\begin{aligned}\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} &= [24.3571 \quad 0.5079]\begin{bmatrix} 1{,}110 \\ 205{,}500 \end{bmatrix} \\ &= 131{,}409.831 \\ \mathbf{y}'\mathbf{y} &= 132{,}100\end{aligned}$$

and

$$n\bar{Y}^2 = 123{,}210$$

Plugging these values into Eq. (C.4.2), we see that $R^2 = 0.9224$, which is about the same as obtained before, save for the rounding errors.

C.5 The Correlation Matrix

In the previous chapters we came across the zero-order, or simple, correlation coefficients r_{12}, r_{13}, r_{23}, and the partial, or first-order, correlations $r_{12.3}$, $r_{13.2}$, $r_{23.1}$, and their interrelationships. In the k-variable case, we shall have in all $k(k-1)/2$ zero-order correlation coefficients. (Why?) These $k(k-1)/2$ correlations can be put into a matrix, called the **correlation matrix *R*** as follows:

$$
\begin{aligned}
R &= \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k1} & r_{k2} & r_{k3} & \cdots & r_{kk} \end{bmatrix} \\
&= \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & 1 & r_{23} & \cdots & r_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{k1} & r_{k2} & r_{k3} & \cdots & 1 \end{bmatrix}
\end{aligned}
\tag{C.5.1}
$$

where the subscript 1, as before, denotes the dependent variable Y (r_{12} means correlation coefficient between Y and X_2, and so on) and where use is made of the fact that the coefficient of correlation of a variable with respect to itself is always 1 ($r_{11} = r_{22} = \cdots = r_{kk} = 1$).

From the correlation matrix $\boldsymbol{R}$ one can obtain correlation coefficients of first order (see Chapter 7) and of higher order such as $r_{12.34\ldots k}$. (See Exercise C.4.) Many computer programs routinely compute the $\boldsymbol{R}$ matrix. We have used the correlation matrix in Chapter 10.

C.6 Hypothesis Testing about Individual Regression Coefficients in Matrix Notation

For reasons spelled out in the previous chapters, if our objective is inference as well as estimation, we shall have to assume that the disturbances u_i follow some probability distribution. Also for reasons given previously, in regression analysis we usually assume that each u_i follows the normal distribution with zero mean and constant variance σ^2. In matrix notation, we have

$$\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{C.6.1}$$

where $\mathbf{u}$ and $\mathbf{0}$ are $n \times 1$ column vectors and $\mathbf{I}$ is an $n \times n$ identity matrix, $\mathbf{0}$ being the **null vector.**

Given the normality assumption, we know that in two- and three-variable linear regression models (1) the OLS estimators $\hat{\beta}_i$ and the ML estimators $\tilde{\beta}_i$ are identical, but the ML estimator $\tilde{\sigma}^2$ is biased, although this bias can be removed by using the unbiased OLS estimator $\hat{\sigma}^2$; and (2) the OLS estimators $\hat{\beta}_i$ are also normally distributed. Generalizing, in the k-variable case we can show that

$$\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}] \tag{C.6.2}$$

that is, each element of $\hat{\boldsymbol{\beta}}$ is normally distributed with mean equal to the corresponding element of true $\boldsymbol{\beta}$ and the variance given by σ^2 times the appropriate diagonal element of the inverse matrix $(\mathbf{X}'\mathbf{X})^{-1}$.

Since in practice σ^2 is unknown, it is estimated by $\hat{\sigma}^2$. Then by the usual shift to the t distribution, it follows that each element of $\hat{\boldsymbol{\beta}}$ follows the t distribution with $n - k$ df. Symbolically,

$$t = \frac{\hat{\beta}_i - \beta_i}{\text{se}\,(\hat{\beta}_i)} \tag{C.6.3}$$

with $n - k$ df, where $\hat{\beta}_i$ is any element of $\hat{\boldsymbol{\beta}}$.

The t distribution can therefore be used to test hypotheses about the true β_i as well as to establish confidence intervals about it. The actual mechanics have already been illustrated in Chapters 5 and 8. For a fully worked example, see Section C.10.

C.7 Testing the Overall Significance of Regression: Analysis of Variance in Matrix Notation

In Chapter 8 we developed the ANOVA technique (1) to test the overall significance of the estimated regression, that is, to test the null hypothesis that the true (partial) slope coefficients are simultaneously equal to zero, and (2) to assess the incremental contribution of an explanatory variable. The ANOVA technique can be easily extended to the k-variable case. Recall that the ANOVA technique consists of decomposing the TSS into two components: the ESS and the RSS. The matrix expressions for these three sums of squares are already given in Eqs. (C.3.16), (C.3.17), and (C.3.18), respectively. The degrees of freedom associated with these sums of squares are $n - 1$, $k - 1$, and $n - k$, respectively. (Why?) Then, following Chapter 8, Table 8.1, we can set up Table C.2.

Assuming that the disturbances u_i are normally distributed and the null hypothesis is $\beta_2 = \beta_3 = \cdots = \beta_k = 0$, and following Chapter 8, one can show that

$$\boxed{F = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2)/(k-1)}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k)}} \tag{C.7.1}$$

follows the F distribution with $k - 1$ and $n - k$ df.

In Chapter 8 we saw that, under the assumptions stated previously, there is a close relationship between F and R^2, namely,

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{8.4.11}$$

Therefore, the ANOVA Table C.2 can be expressed as Table C.3. One advantage of Table C.3 over Table C.2 is that the entire analysis can be done in terms of R^2; one need not consider the term $(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)$, for it drops out in the F ratio.

TABLE C.2
Matrix Formulation of the ANOVA Table for k-Variable Linear Regression Model

Source of Variation	SS	df	MSS
Due to regression (that is, due to $X_2, X_3, \ldots, X_k$)	$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2$	$k-1$	$\dfrac{\hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{k-1}$
Due to residuals	$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	$n-k$	$\dfrac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{n-k}$
Total	$\mathbf{y}'\mathbf{y} - n\bar{Y}^2$	$n-1$	

TABLE C.3
***k*-Variable ANOVA Table in Matrix Form in Terms of R^2**

Source of Variation	SS	df	MSS
Due to regression (that is, due to $X_2, X_3, \ldots, X_k$)	$R^2(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)$	$k-1$	$\dfrac{R^2(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)}{k-1}$
Due to residuals	$(1-R^2)(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)$	$n-k$	$\dfrac{(1-R^2)(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)}{n-k}$
Total	$\mathbf{y}'\mathbf{y} - n\bar{Y}^2$	$n-1$	

C.8 Testing Linear Restrictions: General *F* Testing Using Matrix Notation

In Section 8.6 we introduced the general F test to test the validity of linear restrictions imposed on one or more parameters of the k-variable linear regression model. The appropriate test was given in (8.6.9) (or its equivalent, Eq. [8.6.10]). The matrix counterpart of (8.6.9) can be easily derived.

Let

$\hat{\mathbf{u}}_{\text{R}}$ = the residual vector from the restricted least-squares regression

$\hat{\mathbf{u}}_{\text{UR}}$ = the residual vector from the unrestricted least-squares regression

Then

$$\hat{\mathbf{u}}_{\text{R}}'\hat{\mathbf{u}}_{\text{R}} = \sum \hat{u}_{\text{R}}^2 = \text{RSS from the restricted regression}$$

$$\hat{\mathbf{u}}_{\text{UR}}'\hat{\mathbf{u}}_{\text{UR}} = \sum \hat{u}_{\text{UR}}^2 = \text{RSS from the unrestricted regression}$$

m = number of linear restrictions

k = number of parameters (including the intercept) in the unrestricted regression

n = number of observations

The matrix counterpart of Eq. (8.6.9) is then

$$F = \frac{(\hat{\mathbf{u}}_{\text{R}}'\hat{\mathbf{u}}_{\text{R}} - \hat{\mathbf{u}}_{\text{UR}}'\hat{\mathbf{u}}_{\text{UR}})/m}{(\hat{\mathbf{u}}_{\text{UR}}'\hat{\mathbf{u}}_{\text{UR}})/(n-k)} \tag{C.8.1}$$

which follows the F distribution with $(m, n-k)$ df. As usual, if the computed F value from Eq. (C.8.1) exceeds the critical F value, we can reject the restricted regression; otherwise, we do not reject it.

C.9 Prediction Using Multiple Regression: Matrix Formulation

In Section 8.8 we discussed, using scalar notation, how the estimated multiple regression can be used for predicting (1) the mean and (2) individual values of Y, given the values of the X regressors. In this section we show how to express these predictions in matrix form. We also present the formulas to estimate the variances and standard errors of the predicted values; in Chapter 8 we noted that these formulas are better handled in matrix notation, for the scalar or algebraic expressions of these formulas become rather unwieldy.

Mean Prediction

Let

$$\mathbf{X_0} = \begin{bmatrix} 1 \\ X_{02} \\ X_{03} \\ \vdots \\ X_{0k} \end{bmatrix} \tag{C.9.1}$$

be the vector of values of the X variables for which we wish to predict $\hat{Y}_0$, the mean prediction of Y.

Now the estimated multiple regression, in scalar form, is

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + u_i \tag{C.9.2}$$

which in matrix notation can be written compactly as

$$\hat{Y}_i = \mathbf{x}_i' \hat{\boldsymbol{\beta}} \tag{C.9.3}$$

where $\mathbf{x}_i' = [1 \ X_{2i} \ X_{3i} \ \cdots \ X_{ki}]$ and

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

Equation (C.9.2) or (C.9.3) is of course the mean prediction of Y_i corresponding to given $\mathbf{x}_i'$.

If $\mathbf{x}_i'$ is as given in Eq. (C.9.1), Eq. (C.9.3) becomes

$$(\hat{Y}_i \mid \mathbf{x}_0') = \mathbf{x}_0' \hat{\boldsymbol{\beta}} \tag{C.9.4}$$

where, of course, the values of $\mathbf{x}_0$ are specified. Note that Eq. (C.9.4) gives an unbiased prediction of $E(Y_i \mid \mathbf{x}_0')$, since $E(x_0' \hat{\boldsymbol{\beta}}) = \mathbf{x}_0' \hat{\boldsymbol{\beta}}$. (Why?)

Variance of Mean Prediction

The formula to estimate the variance of $(\hat{Y}_0 \mid \mathbf{x}_0')$ is as follows:[7]

$$\operatorname{var}(\hat{Y}_0 \mid \mathbf{x}_0') = \sigma^2 \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0 \tag{C.9.5}$$

where σ^2 is the variance of u_i, $\mathbf{x}_0'$ are the given values of the X variables for which we wish to predict, and $(\mathbf{X}'\mathbf{X})$ is the matrix given in Eq. (C.3.9). In practice, we replace σ^2 by its unbiased estimator $\hat{\sigma}^2$.

We will illustrate mean prediction and its variance in the next section.

Individual Prediction

As pointed out in Chapters 5 and 8, the individual prediction of $Y \ (= Y_0)$ is also given by Eq. (C.9.3) or more specifically by Eq. (C.9.4). The difference between mean and individual predictions lies in their variances.

Variance of Individual Prediction

The formula for the variance of an individual prediction is as follows:[8]

$$\operatorname{var}(Y_0 \mid \mathbf{x}_0) = \sigma^2 [1 + \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0] \tag{C.9.6}$$

where $\operatorname{var}(Y_0 \mid \mathbf{x}_0)$ stands for $E[Y_0 - \hat{Y}_0 \mid X]^2$. In practice we replace σ^2 by its unbiased estimator $\hat{\sigma}^2$. We illustrate this formula in the next section.

[7] For derivation, see J. Johnston, *Econometrics Methods,* McGraw-Hill, 3d ed., New York, 1984, pp. 195–196.

[8] Ibid.

C.10 Summary of the Matrix Approach: An Illustrative Example

Consider the data given in Table C.4. These data pertain to per capita personal consumption expenditure (PPCE) and per capital personal disposable income (PPDI) and time or the trend variable. By including the trend variable in the model, we are trying to find out the relationship of PPCE to PPDI net of the trend variable (which may represent a host of other factors, such as technology, change in tastes, etc.).

For empirical purposes, therefore, the regression model is

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \tag{C.10.1}$$

where Y = per capita consumption expenditure, X_2 = per capita disposable income, and X_3 = time. The data required to run the regression (C.10.1) are given in Table C.4.

In matrix notation, our problem may be shown as follows:

$$\underset{15\times 1}{\begin{bmatrix} 1673 \\ 1688 \\ 1666 \\ 1735 \\ 1749 \\ 1756 \\ 1815 \\ 1867 \\ 1948 \\ 2048 \\ 2128 \\ 2165 \\ 2257 \\ 2316 \\ 2324 \end{bmatrix}} = \underset{15\times 3}{\begin{bmatrix} 1 & 1839 & 1 \\ 1 & 1844 & 2 \\ 1 & 1831 & 3 \\ 1 & 1881 & 4 \\ 1 & 1883 & 5 \\ 1 & 1910 & 6 \\ 1 & 1969 & 7 \\ 1 & 2016 & 8 \\ 1 & 2126 & 9 \\ 1 & 2239 & 10 \\ 1 & 2336 & 11 \\ 1 & 2404 & 12 \\ 1 & 2487 & 13 \\ 1 & 2535 & 14 \\ 1 & 2595 & 15 \end{bmatrix}} \underset{3\times 1}{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}} + \underset{15\times 1}{\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \\ \hat{u}_8 \\ \hat{u}_9 \\ \hat{u}_{10} \\ \hat{u}_{11} \\ \hat{u}_{12} \\ \hat{u}_{13} \\ \hat{u}_{14} \\ \hat{u}_{15} \end{bmatrix}} \tag{C.10.2}$$

$$\underset{15\times 1}{\mathbf{y}} = \underset{15\times 3}{\mathbf{X}} \quad \underset{3\times 1}{\hat{\boldsymbol{\beta}}} + \underset{15\times 1}{\hat{\mathbf{u}}}$$

TABLE C.4
Per Capita Personal Consumption Expenditure (PPCE) and Per Capita Personal Disposable Income (PPDI) in the United States, 1956–1970, in 1958 Dollars

PPCE, Y	PPDI, X_2	Time, X_3	PPCE, Y	PPDI, X_2	Time, X_3
1673	1839	1 (= 1956)	1948	2126	9
1688	1844	2	2048	2239	10
1666	1831	3	2128	2336	11
1735	1881	4	2165	2404	12
1749	1883	5	2257	2487	13
1756	1910	6	2316	2535	14
1815	1969	7	2324	2595	15 (= 1970)
1867	2016	8			

Source: *Economic Report of the President,* January 1972, Table B-16.

From the preceding data we obtain the following quantities:

$$\bar{Y} = 1942.333 \qquad \bar{X}_2 = 2126.333 \qquad \bar{X}_3 = 8.0$$

$$\sum (Y_i - \bar{Y})^2 = 830{,}121.333$$

$$\sum (X_{2i} - \bar{X}_2)^2 = 1{,}103{,}111.333 \qquad \sum (X_{3i} - \bar{X}_3)^2 = 280.0$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2n} \\ X_{31} & X_{32} & X_{33} & \cdots & X_{3n} \end{bmatrix} \begin{bmatrix} 1 & X_{21} & X_{31} \\ 1 & X_{22} & X_{32} \\ 1 & X_{23} & X_{33} \\ \vdots & \vdots & \vdots \\ 1 & X_{2n} & X_{3n} \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{2i}X_{3i} & \sum X_{3i}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 15 & 31{,}895 & 120 \\ 31{,}895 & 68{,}922.513 & 272{,}144 \\ 120 & 272{,}144 & 1240 \end{bmatrix} \qquad \textbf{(C.10.3)}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 29{,}135 \\ 62{,}905{,}821 \\ 247{,}934 \end{bmatrix} \qquad \textbf{(C.10.4)}$$

Using the rules of matrix inversion given in **Appendix B,** one can see that

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 37.232491 & -0.0225082 & 1.336707 \\ -0.0225082 & 0.0000137 & -0.0008319 \\ 1.336707 & -0.0008319 & 0.054034 \end{bmatrix} \qquad \textbf{(C.10.5)}$$

Therefore,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 300.28625 \\ 0.74198 \\ 8.04356 \end{bmatrix} \qquad \textbf{(C.10.6)}$$

The residual sum of squares can now be computed as

$$\begin{aligned} \sum \hat{u}_i^2 &= \hat{\mathbf{u}}'\hat{\mathbf{u}} \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \\ &= 57{,}420{,}003 - [300.28625 \quad 0.74198 \quad 8.04356] \begin{bmatrix} 29{,}135 \\ 62{,}905{,}821 \\ 247{,}934 \end{bmatrix} \\ &= 1976.85574 \end{aligned} \qquad \textbf{(C.10.7)}$$

whence we obtain

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{12} = 164.73797 \qquad \textbf{(C.10.8)}$$

The variance-covariance matrix for $\hat{\boldsymbol{\beta}}$ can therefore be shown as

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 6133.650 & -3.70794 & 220.20634 \\ -3.70794 & 0.00226 & -0.13705 \\ 220.20634 & -0.13705 & 8.90155 \end{bmatrix} \qquad \textbf{(C.10.9)}$$

The diagonal elements of this matrix give the variances of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, respectively, and their positive square roots give the corresponding standard errors.

From the previous data, it can be readily verified that

$$\text{ESS: } \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 = 828{,}144.47786 \tag{C.10.10}$$

$$\text{TSS: } \mathbf{y}'\mathbf{y} - n\bar{Y}^2 = 830{,}121.333 \tag{C.10.11}$$

Therefore,

$$\begin{aligned} R^2 &= \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} \\ &= \frac{828{,}144.47786}{830{,}121.333} \\ &= 0.99761 \end{aligned} \tag{C.10.12}$$

Applying Eq. (7.8.4) the **adjusted coefficient of determination** can be seen to be

$$\bar{R}^2 = 0.99722 \tag{C.10.13}$$

Collecting our results thus far, we have

$$\begin{aligned} \hat{Y}_i &= 300.28625 \quad + \ 0.74198X_{2i} \ + 8.04356X_{3i} \\ &\quad\ (78.31763) \quad (0.04753) \qquad (2.98354) \\ t &= \quad (3.83421) \quad (15.60956) \qquad (2.69598) \\ &\quad R^2 = 0.99761 \qquad \bar{R}^2 = 0.99722 \qquad \text{df} = 12 \end{aligned} \tag{C.10.14}$$

The interpretation of Eq. (C.10.14) is this: If both X_2 and X_3 are fixed at zero value, the average value of per capita personal consumption expenditure is estimated at about \$300. As usual, this mechanical interpretation of the intercept should be taken with a grain of salt. The partial regression coefficient of 0.74198 means that, holding all other variables constant, an increase in per capita income of, say, a dollar is accompanied by an increase in the mean per capita personal consumption expenditure of about 74 cents. In short, the marginal propensity to consume is estimated to be about 0.74, or 74 percent. Similarly, holding all other variables constant, the mean per capita personal consumption expenditure increased at the rate of about \$8 per year during the period of the study, 1956–1970. The R^2 value of 0.9976 shows that the two explanatory variables accounted for over 99 percent of the variation in per capita consumption expenditure in the United States over the period 1956–1970. Although $\bar{R}^2$ dips slightly, it is still very high.

Turning to the statistical significance of the estimated coefficients, we see from Eq. (C.10.14) that each of the estimated coefficients is *individually* statistically significant at, say, the 5 percent level of significance: The ratios of the estimated coefficients to their standard errors (that is, t ratios) are 3.83421, 15.61077, and 2.69598, respectively. Using a two-tail t test at the 5 percent level of significance, we see that the critical t value for 12 df is 2.179. Each of the computed t values exceeds this critical value. Hence, individually we may reject the null hypothesis that the true population value of the relevant coefficient is zero.

As noted previously, we cannot apply the usual t test to test the hypothesis that $\beta_2 = \beta_3 = 0$ simultaneously because the t-test procedure assumes that an independent sample is drawn every time the t test is applied. If the same sample is used to test hypotheses about β_2 and β_3 simultaneously, it is likely that the estimators $\hat{\beta}_2$ and $\hat{\beta}_3$ are correlated, thus

TABLE C.5
The ANOVA Table for the Data of Table C.4

Source of Variation	SS	df	MSS
Due to X_2, X_3	828,144.47786	2	414,072.3893
Due to residuals	1,976.85574	12	164.73797
Total	830,121.33360	14	

violating the assumption underlying the t-test procedure.[9] As a matter of fact, a look at the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ given in Eq. (C.10.9) shows that the estimators $\hat{\beta}_2$ and $\hat{\beta}_3$ are negatively correlated (the covariance between the two is -0.13705). Hence we cannot use the t test to test the null hypothesis that $\beta_2 = \beta_3 = 0$.

But recall that a null hypothesis like $\beta_2 = \beta_3 = 0$, simultaneously, can be tested by the analysis of variance technique and the attendant F test, which were introduced in Chapter 8. For our problem, the analysis of variance table is Table C.5. Under the usual assumptions, we obtain

$$F = \frac{414{,}072.3893}{164.73797} = 2513.52 \tag{C.10.15}$$

which is distributed as the F distribution with 2 and 12 df. The computed F value is obviously highly significant; we can reject the null hypothesis that $\beta_2 = \beta_3 = 0$, that is, that per capita personal consumption expenditure is not linearly related to per capita disposable income and trend.

In Section C.9 we discussed the mechanics of forecasting, mean as well as individual. Assume that for 1971 the PPDI figure is \$2,610 and we wish to forecast the PPCE corresponding to this figure. Then, the mean as well as individual forecast of PPCE for 1971 is the same and is given as

$$\begin{aligned}
(\text{PPCE}_{1971} \mid \text{PPDI}_{1971}, X_3 = 16) &= \mathbf{x}'_{1971}\hat{\boldsymbol{\beta}} \\
&= [1 \quad 2610 \quad 16] \begin{bmatrix} 300.28625 \\ 0.74198 \\ 8.04356 \end{bmatrix} \\
&= 2365.55
\end{aligned} \tag{C.10.16}$$

where use is made of Eq. (C.9.3).

The variances of $\hat{Y}_{1971}$ and Y_{1971}, as we know from Section C.9, are different and are as follows:

$$\begin{aligned}
\operatorname{var}(\hat{Y}_{1971} \mid \mathbf{x}'_{1971}) &= \hat{\sigma}^2[\mathbf{x}'_{1971}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{1971}] \\
&= 164.73797[1 \quad 2610 \quad 16](\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} 1 \\ 2610 \\ 16 \end{bmatrix}
\end{aligned} \tag{C.10.17}$$

where $(\mathbf{X}'\mathbf{X})^{-1}$ is as shown in Eq. (C.10.5). Substituting this into Eq. (C.10.17), the reader should verify that

$$\operatorname{var}(\hat{Y}_{1971} \mid \mathbf{x}'_{1971}) = 48.6426 \tag{C.10.18}$$

[9]See Section 8.4 for details.

and therefore

$$\text{se}\,(\hat{Y}_{1971} \mid \mathbf{x}'_{1971}) = 6.9744$$

We leave it to the reader to verify, using Eq. (C.9.6), that

$$\text{var}\,(Y_{1971} \mid \mathbf{x}'_{1971}) = 213.3806 \tag{C.10.19}$$

and

$$\text{se}\,(Y_{1971} \mid \mathbf{x}'_{1971}) = 14.6076$$

Note: $\text{var}\,(Y_{1971} \mid \mathbf{x}'_{1971}) = E[Y_{1971} - \hat{Y}_{1971} \mid \mathbf{x}'_{1971}]^2$.

In Section C.5 we introduced the correlation matrix $\boldsymbol{R}$. For our data, the correlation matrix is as follows:

$$R = \begin{array}{c} \\ Y \\ X_2 \\ X_3 \end{array} \begin{array}{c} \begin{array}{ccc} Y & X_2 & X_3 \end{array} \\ \begin{bmatrix} 1 & 0.9980 & 0.9743 \\ 0.9980 & 1 & 0.9664 \\ 0.9743 & 0.9664 & 1 \end{bmatrix} \end{array} \tag{C.10.20}$$

Note that in Eq. (C.10.20) we have bordered the correlation matrix by the variables of the model so that we can readily identify which variables are involved in the computation of the correlation coefficient. Thus, the coefficient 0.9980 in the first row of matrix (C.10.20) tells us that it is the correlation coefficient between Y and X_2 (that is, r_{12}). From the zero-order correlations given in the correlation matrix (C.10.20) one can easily derive the first-order correlation coefficients. (See Exercise C.7.)

C.11 Generalized Least Squares (GLS)

On several occasions we have mentioned that OLS is a special case of GLS. To see this, return to Eq. (C.2.2). To take into account heteroscedastic variances (the elements on the main diagonal of Eq. [C.2.2]) and autocorrelations in the error terms (the elements off the main diagonal of Eq. [C.2.2]), assume that

$$\mathrm{E}(\mathbf{uu}') = \sigma^2\mathbf{V} \tag{C.11.1}$$

where $\mathbf{V}$ is a known $n \times n$ matrix.

Therefore, if our model is:

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

where $E(\mathbf{u}) = \mathbf{0}$ and var-cov $(\mathbf{u}) = \sigma^2\mathbf{V}$. In case σ^2 is unknown, which is typically the case, $\mathbf{V}$ then represents the assumed structure of variances and covariances among the random errors u_t.

Under the stated condition of the variance-covariance of the error terms, it can be shown that

$$\boldsymbol{\beta}^{\text{gls}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{C.11.2}$$

$\boldsymbol{\beta}^{\text{gls}}$ is known as the **generalized least-squares (GLS) estimator** of $\boldsymbol{\beta}$.

It can also be shown that

$$\text{var-cov}\,(\boldsymbol{\beta}^{\text{gls}}) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \tag{C.11.3}$$

It can be proved that $\boldsymbol{\beta}^{\text{gls}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$.

If it is assumed that the variance of each error term is the same constant σ^2 and the error terms are mutually uncorrelated, then the **V** matrix reduces to the identity matrix, as shown in Eq. (C.2.3). If the error terms are mutually uncorrelated but they have different (i.e., heteroscedastic) variances, then the **V** matrix will be diagonal with the unequal variances along the main diagonal. Of course, if there is heteroscedasticity as well as autocorrelation, then the **V** matrix will have entries on the main diagonal as well as on the off diagonal.

The real problem in practice is that we do not know σ^2 as well as the true variances and covariances (i.e., the structure of the **V** matrix). As a solution, we can use the method of **estimated (or feasible) generalized least squares (EGLS).** Here we first estimate our model by OLS, disregarding the problems of heteroscedasticity and/or autocorrelation. We obtain the residuals from this model and form the (estimated) variance-covariance matrix of the error term by replacing the entries in the expression just before Eq. (C.2.2) by the estimated u, namely, $\hat{u}$. It can be shown that EGLS estimators are consistent estimators of GLS. Symbolically,

$$\boldsymbol{\beta}^{\text{egls}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}) \tag{C.11.4}$$

$$\text{var-cov}\,(\boldsymbol{\beta}^{\text{egls}}) = \sigma^2(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \tag{C.11.5}$$

where $\hat{\mathbf{V}}$ is an estimate of **V**.

C.12 Summary and Conclusions

The primary purpose of this appendix was to introduce the matrix approach to the classical linear regression model. Although very few new concepts of regression analysis were introduced, the matrix notation provides a compact method of dealing with linear regression models involving any number of variables.

In concluding this appendix, note that if the Y and X variables are measured in the deviation form, that is, as deviations from their sample means, there are a few changes in the formulas presented previously. These changes are listed in Table C.6.[10] As this table shows, in

TABLE C.6
k*-Variable Regression Model in Original Units and in the Deviation Form

Original Units		Deviation Form	
$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$	(C.3.2)	$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$ The column of 1's in the **X** matrix drops out. (Why?)	
$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$	(C.3.11)	Same	
$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$	(C.3.13)	Same	
$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	(C.3.18)	Same	
$\sum y_i^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2$	(C.3.16)	$\sum y_i^2 = \mathbf{y}'\mathbf{y}$	(C.12.1)
$\text{ESS} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2$	(C.3.17)	$\text{ESS} = \boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$	(C.12.2)
$R^2 = \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2}$	(C.4.2)	$R^2 = \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}$	(C.12.3)

*Note that although in both cases the symbols for the matrices and vectors are the same, in the deviation form the elements of the matrices and vectors are assumed to be deviations rather than the raw data. Note also that in the deviation form $\hat{\boldsymbol{\beta}}$ is of order $k-1$ and the var-cov $(\hat{\boldsymbol{\beta}})$ is of order $(k-1)(k-1)$.

[10] In these days of high-speed computers there may not be need for the deviation form. But it simplifies formulas and therefore calculations if one is working with a desk calculator and dealing with large numbers.

the deviation form the correction for mean $n\bar{Y}^2$ drops out from the TSS and ESS. (Why?) This loss results in a change for the formula for R^2. Otherwise, most of the formulas developed in the original units of measurement hold true for the deviation form.

EXERCISES

C.1. For the illustrative example discussed in Section C.10 the $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$ using the data in the deviation form are as follows:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1{,}103{,}111.333 & 16{,}984 \\ 16{,}984 & 280 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 955{,}099.333 \\ 14{,}854.000 \end{bmatrix}$$

a. Estimate β_2 and β_3.

b. How would you estimate β_1?

c. Obtain the variance of $\hat{\beta}_2$ and $\hat{\beta}_3$ and their covariances.

d. Obtain R^2 and $\bar{R}^2$.

e. Comparing your results with those given in Section C.10, what do you find are the advantages of the deviation form?

C.2. Refer to Exercise 22.23. Using the data given therein, set up the appropriate $(\mathbf{X}'\mathbf{X})$ matrix and the $\mathbf{X}'\mathbf{y}$ vector and estimate the parameter vector $\boldsymbol{\beta}$ and its variance-covariance matrix. Also obtain R^2. How would you test the hypothesis that the elasticities of M1 with respect to GDP and interest rate R are numerically the same?

C.3. *Testing the equality of two regression coefficients.* Suppose that you are given the following regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

and you want to test the hypothesis that $\beta_2 = \beta_3$. If we assume that the u_i are normally distributed, it can be shown that

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\sqrt{\operatorname{var}(\hat{\beta}_2) + \operatorname{var}(\hat{\beta}_3) - 2\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3)}}$$

follows the t distribution with $n - 3$ df (see Section 8.5). (In general, for the k-variable case the df are $n - k$.) Therefore, the preceding t test can be used to test the null hypothesis $\beta_2 = \beta_3$.

Apply the preceding t test to test the hypothesis that the true values of β_2 and β_3 in the regression (C.10.14) are identical.

Hint: Use the var-cov matrix of $\boldsymbol{\beta}$ given in Eq. (C.10.9).

C.4. *Expressing higher-order correlations in terms of lower-order correlations.* Correlation coefficients of order p can be expressed in terms of correlation coefficients of order $p - 1$ by the following **reduction formula:**

$$r_{12.345\ldots p} = \frac{r_{12.345\ldots(p-1)} - [r_{1p.345\ldots(p-1)} r_{2p.345\ldots(p-1)}]}{\sqrt{[1 - r^2_{1p.345\ldots(p-1)}]}\sqrt{[1 - r^2_{2p.345\ldots(p-1)}]}}$$

Thus,

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

as found in Chapter 7.

You are given the following correlation matrix:

$$\mathbf{R} = \begin{array}{c} \\ Y \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{bmatrix} 1 & 0.44 & -0.34 & -0.31 & -0.14 \\ & 1 & 0.25 & -0.19 & -0.35 \\ & & 1 & 0.44 & 0.33 \\ & & & 1 & 0.85 \\ & & & & 1 \end{bmatrix} \end{array}$$

Find the following:

a. $r_{12.345}$ *b.* $r_{12.34}$ *c.* $r_{12.3}$

d. $r_{13.245}$ *e.* $r_{13.24}$ *f.* $r_{13.2}$

C.5. *Expressing higher-order regression coefficients in terms of lower-order regression coefficients.* A regression coefficient of order p can be expressed in terms of a regression coefficient of order $p - 1$ by the following reduction formula:

$$\hat{\beta}_{12.345\ldots p} = \frac{\hat{\beta}_{12.345\ldots(p-1)} - \left[\hat{\beta}_{1p.345\ldots(p-1)}\hat{\beta}_{p2.345\ldots(p-1)}\right]}{1 - \hat{\beta}_{2p.345\ldots(p-1)}\hat{\beta}_{p2.345\ldots(p-1)}}$$

Thus,

$$\hat{\beta}_{12.3} = \frac{\hat{\beta}_{12} - \hat{\beta}_{13}\hat{\beta}_{32}}{1 - \hat{\beta}_{23}\hat{\beta}_{32}}$$

where $\beta_{12.3}$ is the slope coefficient in the regression of y on X_2 holding X_3 constant. Similarly, $\beta_{12.34}$ is the slope coefficient in the regression of Y on X_2 holding X_3 and X_4 constant, and so on.

Using the preceding formula, find expressions for the following regression coefficients in terms of lower-order regression coefficients: $\hat{\beta}_{12.3456}$, $\hat{\beta}_{12.345}$, and $\hat{\beta}_{12.34}$.

C.6. Establish the following identity:

$$\hat{\beta}_{12.3}\hat{\beta}_{23.1}\hat{\beta}_{31.2} = r_{12.3}r_{23.1}r_{31.2}$$

C.7. For the correlation matrix $\boldsymbol{R}$ given in Eq. (C.10.20) find all the first-order partial correlation coefficients.

C.8. In studying the variation in crime rates in certain large cities in the United States, Ogburn obtained the following data:*

$$\begin{array}{ll} \bar{Y} = 19.9 & S_1 = 7.9 \\ \bar{X}_2 = 49.2 & S_2 = 1.3 \\ \bar{X}_3 = 10.2 & S_3 = 4.6 \\ \bar{X}_4 = 481.4 & S_4 = 74.4 \\ \bar{X}_5 = 41.6 & S_5 = 10.8 \end{array} \qquad \boldsymbol{R} = \begin{array}{c} \\ Y \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} \begin{array}{c} \begin{array}{ccccc} Y & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{bmatrix} 1 & 0.44 & -0.34 & -0.31 & -0.14 \\ & 1 & 0.25 & -0.19 & -0.35 \\ & & 1 & 0.44 & 0.33 \\ & & & 1 & 0.85 \\ & & & & 1 \end{bmatrix} \end{array}$$

*W. F. Ogburn, "Factors in the Variation of Crime among Cities," *Journal of American Statistical Association,* vol. 30, 1935, p. 12.

where Y = crime rate, number of known offenses per thousand of population

X_2 = percentage of male inhabitants

X_3 = percentage of total inhabitants who are foreign-born males

X_4 = number of children under 5 years of age per thousand married women between ages 15 and 44 years

X_5 = church membership, number of church members 13 years of age and over per 100 of total population 13 years of age and over; S_1 to S_5 are the sample standard deviations of variables Y through X_5 and $\boldsymbol{R}$ is the correlation matrix

a. Treating Y as the dependent variable, obtain the regression of Y on the four X variables and interpret the estimated regression.

b. Obtain $r_{1\,2.3}$, $r_{1\,4.3\,5}$, and $r_{1\,5.3\,4}$.

c. Obtain R^2 and test the hypothesis that all partial slope coefficients are simultaneously equal to zero.

C.9. The following table gives data on output and total cost of production of a commodity in the short run. (See Example 7.4.)

Output	Total Cost, $
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

To test whether the preceding data suggest the U-shaped average and marginal cost curves typically encountered in the short run, one can use the following model:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i$$

where Y = total cost and X = output. The additional explanatory variables X_i^2 and X_i^3 are derived from X.

a. Express the data in the deviation form and obtain $(\mathbf{X}'\mathbf{X})$, $(\mathbf{X}'\mathbf{y})$, and $(\mathbf{X}'\mathbf{X})^{-1}$.

b. Estimate β_2, β_3, and β_4.

c. Estimate the var-cov matrix of $\hat{\boldsymbol{\beta}}$.

d. Estimate β_1. Interpret $\hat{\beta}_1$ in the context of the problem.

e. Obtain R^2 and $\bar{R}^2$.

f. A priori, what are the signs of β_2, β_3, and β_4? Why?

g. From the total cost function given previously obtain expressions for the marginal and average cost functions.

h. Fit the average and marginal cost functions to the data and comment on the fit.

i. If $\beta_3 = \beta_4 = 0$, what is the nature of the marginal cost function? How would you test the hypothesis that $\beta_3 = \beta_4 = 0$?

j. How would you derive the total variable cost and average variable cost functions from the given data?

TABLE C.7
Labor Force Participation Experience of the Urban Poor: Census Tracts, New York City, 1970

Source: Census Tracts: New York, Bureau of the Census, U.S. Department of Commerce, 1970.

Tract No.	% in Labor Force, Y*	Mean Family Income, X_2†	Mean Family Size, X_3	Unemployment Rate, X_4‡
137	64.3	1,998	2.95	4.4
139	45.4	1,114	3.40	3.4
141	26.6	1,942	3.72	1.1
142	87.5	1,998	4.43	3.1
143	71.3	2,026	3.82	7.7
145	82.4	1,853	3.90	5.0
147	26.3	1,666	3.32	6.2
149	61.6	1,434	3.80	5.4
151	52.9	1,513	3.49	12.2
153	64.7	2,008	3.85	4.8
155	64.9	1,704	4.69	2.9
157	70.5	1,525	3.89	4.8
159	87.2	1,842	3.53	3.9
161	81.2	1,735	4.96	7.2
163	67.9	1,639	3.68	3.6

*Y = family heads under 65 years old.
†X_2 = dollars.
‡X_4 = percent of civilian labor force unemployed.

C.10. In order to study the labor force participation of urban poor families (families earning less than \$3,943 in 1969), the data in Table C.7 were obtained from the 1970 Census of Population.

a. Using the regression model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$, obtain the estimates of the regression coefficients and interpret your results.

b. A priori, what are the expected signs of the regression coefficients in the preceding model and why?

c. How would you test the hypothesis that the overall unemployment rate has no effect on the labor force participation of the urban poor in the census tracts given in the accompanying table?

d. Should any variables be dropped from the preceding model? Why?

e. What other variables would you consider for inclusion in the model?

C.11. In an application of the Cobb–Douglas production function the following results were obtained:

$$\widehat{\ln Y}_i = 2.3542 + \underset{(0.3022)}{0.9576} \ln X_{2i} + \underset{(0.3571)}{0.8242} \ln X_{3i}$$
$$R^2 = 0.8432 \qquad \text{df} = 12$$

where Y = output, X_2 = labor input, and X_3 = capital input, and where the figures in parentheses are the estimated standard errors.

a. As noted in Chapter 7, the coefficients of the labor and capital inputs in the preceding equation give the elasticities of output with respect to labor and capital. Test the hypothesis that these elasticities are *individually* equal to unity.

b. Test the hypothesis that the labor and capital elasticities are equal, assuming (*i*) the covariance between the estimated labor and capital coefficients is zero, and (*ii*) it is −0.0972.

c. How would you test the overall significance of the preceding regression equation?

*C.12. Express the likelihood function for the k-variable regression model in matrix notation and show that $\tilde{\boldsymbol{\beta}}$, the vector of maximum likelihood estimators, is identical to $\hat{\boldsymbol{\beta}}$, the vector of OLS estimators of the k-variable regression model.

C.13. *Regression using standardized variables.* Consider the following sample regression functions (SRFs):

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \tag{1}$$

$$Y_i^* = b_1 + b_2 X_{2i}^* + b_3 X_{3i}^* + \hat{u}_i^* \tag{2}$$

where

$$Y_i^* = \frac{Y_i - \bar{Y}}{s_Y}$$

$$X_{2i}^* = \frac{X_{2i} - \bar{X}_2}{s_2}$$

$$X_{3i}^* = \frac{X_{3i} - \bar{X}_3}{s_3}$$

where the s's denote the sample standard deviations. As noted in Chapter 6, Section 6.3, the starred variables above are known as the *standardized variables*. These variables have zero means and unit $(= 1)$ standard deviations. Expressing all the variables in the deviation form, show the following for model (2):

a. $\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & r_{23} \\ r_{23} & 1 \end{bmatrix} n$

b. $\mathbf{X}'\mathbf{y} = \begin{bmatrix} r_{12} \\ r_{13} \end{bmatrix} n$

c. $\mathbf{X}'\mathbf{X}^{-1} = \dfrac{1}{n\left(1 - r_{23}^2\right)} \begin{bmatrix} 1 & -r_{23} \\ -r_{23} & 1 \end{bmatrix}$

d. $\hat{\boldsymbol{\beta}} = \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \dfrac{1}{1 - r_{23}^2} \begin{bmatrix} r_{12} - r_{23} r_{13} \\ r_{13} - r_{23} r_{12} \end{bmatrix}$

e. $b_1 = 0$

Also establish the relationship between the b's and the $\hat{\beta}$'s.

(Note that in the preceding relations n denotes the sample size; r_{12}, r_{13}, and r_{23} denote the correlations between Y and X_2, between Y and X_3, and between X_2 and X_3, respectively.)

C.14. Verify Eqs. (C.10.18) and (C.10.19).

*C.15. *Constrained least-squares.* Assume

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{1}$$

which we want to estimate subject to a set of equality restrictions or constraints:

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r} \tag{2}$$

where $\mathbf{R}$ is a *known* matrix of order qxk $(q \leq k)$ and $\mathbf{r}$ is a *known* vector of q elements. To illustrate, suppose our model is

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i \tag{3}$$

*Optional.

and suppose we want to estimate this model subject to these restrictions:

$$\begin{aligned} \beta_2 - \beta_3 &= 0 \\ \beta_4 + \beta_5 &= 1 \end{aligned} \tag{4}$$

We can use some of the techniques discussed in Chapter 8 to incorporate these restrictions (e.g., $\beta_2 = \beta_3$ and $\beta_4 = 1 - \beta_5$, thus removing β_2 and β_4 from the model) and test for the validity of these restrictions by the F test discussed there. But a more direct way of estimating Eq. (3) incorporating the restrictions (4) directly in the estimating procedure is first to express the restrictions in the form of Eq. (2), which in the present case becomes

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad \mathbf{r} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{5}$$

Letting $\boldsymbol{\beta}^*$ denote the restricted least-squares or constrained least-squares estimator, one can show that $\boldsymbol{\beta}^*$ can be estimated by the following formula:*

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}) \tag{6}$$

where $\hat{\boldsymbol{\beta}}$ is the usual (unconstrained) estimator estimated from the usual formula $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

a. What is the β vector in Eq. (3)?

b. Given this β vector, verify that the **R** matrix and **r** vector given in Eq. (5) do in fact incorporate the restrictions in Eq. (4).

c. Write down the **R** and **r** in the following cases:

(*i*) $\beta_2 = \beta_3 = \beta_4 = 2$

(*ii*) $\beta_2 = \beta_3$ and $\beta_4 = \beta_5$

(*iii*) $\beta_2 - 3\beta_3 = 5\beta_4$

(*iv*) $\beta_2 + 3\beta_3 = 0$

d. When will $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}$?

Appendix CA

CA.1 Derivation of *k* Normal or Simultaneous Equations

Differentiating

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2$$

partially with respect to $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$, we obtain

$$\begin{aligned}
\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} &= 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})(-1) \\
\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} &= 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})(-X_{2i}) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_k} &= 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{ki} - \cdots - \hat{\beta}_k X_{ki})(-X_{ki})
\end{aligned}$$

Setting the preceding partial derivatives equal to zero and rearranging the terms, we obtain the k normal equations given in Eq. (C.3.8).

*See J. Johnston, op. cit., p. 205.

CA.2 Matrix Derivation of Normal Equations

From Eq. (C.3.7) we obtain

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

Using the rules of matrix differentiation given in **Appendix B, Section B.6,** we obtain

$$\frac{\partial(\hat{\mathbf{u}}'\hat{\mathbf{u}})}{\partial\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

Setting the preceding equation to zero gives

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

whence $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, provided the inverse exists.

CA.3 Variance–Covariance Matrix of $\hat{\boldsymbol{\beta}}$

From Eq. (C.3.11) we obtain

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

Substituting $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ into the preceding expression gives

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u})\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \qquad (1)\\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\end{aligned}$$

Therefore,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \qquad (2)$$

By definition

$$\begin{aligned}\text{var-cov}\,(\hat{\boldsymbol{\beta}}) &= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']\\ &= E\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]'\} \qquad (3)\\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]\end{aligned}$$

where in the last step use is made of the fact that $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Noting that the X's are nonstochastic, on taking expectation of Eq. (3) we obtain

$$\begin{aligned}\text{var-cov}\,(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

which is the result given in Eq. (C.3.13). Note that in deriving the preceding result use is made of the assumption that $E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$.

CA.4 BLUE Property of OLS Estimators

From Eq. (C.3.11) we have

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

Since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is a matrix of fixed numbers, $\hat{\boldsymbol{\beta}}$ is a linear function of Y. Hence, by definition it is a linear estimator.

Recall that the PRF is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{2}$$

Substituting this into Eq. (1), we obtain

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \tag{3}$$

$$= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \tag{4}$$

since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$.

Taking expectation of Eq. (4) gives

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u})$$

$$= \boldsymbol{\beta} \tag{5}$$

since $E(\boldsymbol{\beta}) = \boldsymbol{\beta}$ (why?) and $E(\mathbf{u}) = \mathbf{0}$ by assumption, which shows that $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$.

Let $\hat{\boldsymbol{\beta}}^*$ be any other linear estimator of $\boldsymbol{\beta}$, which can be written as

$$\boldsymbol{\beta}^* = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}]\mathbf{y} \tag{6}$$

where $\mathbf{C}$ is a matrix of constants.

Substituting for $\mathbf{y}$ from Eq. (2) into Eq. (6), we get

$$\hat{\boldsymbol{\beta}}^* = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}](\mathbf{X}\boldsymbol{\beta} + \mathbf{u})$$

$$= \boldsymbol{\beta} + \mathbf{C}\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{C}\mathbf{u} \tag{7}$$

Now if $\hat{\boldsymbol{\beta}}^*$ is to be an unbiased estimator of $\boldsymbol{\beta}$, we must have

$$\mathbf{C}\mathbf{X} = 0 \qquad \text{(Why?)} \tag{8}$$

Using Eq. (8), Eq. (7) can be written as

$$\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{C}\mathbf{u} \tag{9}$$

By definition, the var-cov $(\hat{\boldsymbol{\beta}}^*)$ is

$$E(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})' = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{C}\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{C}\mathbf{u}]' \tag{10}$$

Making use of the properties of matrix inversion and transposition and after algebraic simplification, we obtain

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}^*) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{C}\mathbf{C}'$$

$$= \text{var-cov}\,(\hat{\boldsymbol{\beta}}) + \sigma^2\mathbf{C}\mathbf{C}' \tag{11}$$

which shows that the variance-covariance matrix of the alternative unbiased linear estimator $\hat{\boldsymbol{\beta}}^*$ is equal to the variance-covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}}$ plus σ^2 times $\mathbf{C}\mathbf{C}'$, which is a positive semidefinite* matrix. Hence the variances of a given element of $\hat{\boldsymbol{\beta}}^*$ must necessarily be equal to or greater than the corresponding element of $\hat{\boldsymbol{\beta}}$, which shows that $\hat{\boldsymbol{\beta}}$ is BLUE. Of course, if $\mathbf{C}$ is a null matrix, i.e., $\mathbf{C} = \mathbf{0}$, then $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}$, which is another way of saying that if we have found a BLUE estimator, it must be the least-squares estimator $\hat{\boldsymbol{\beta}}$.

*See references in **Appendix B.**