7

MULTIPLE REGRESSION ANALYSIS: THE PROBLEM OF ESTIMATION

The two-variable model studied extensively in the previous chapters is often inadequate in practice. In our consumption–income example, for instance, it was assumed implicitly that only income X affects consumption Y. But economic theory is seldom so simple for, besides income, a number of other variables are also likely to affect consumption expenditure. An obvious example is wealth of the consumer. As another example, the demand for a commodity is likely to depend not only on its own price but also on the prices of other competing or complementary goods, income of the consumer, social status, etc. Therefore, we need to extend our simple two-variable regression model to cover models involving more than two variables. Adding more variables leads us to the discussion of multiple regression models, that is, models in which the dependent variable, or regressand, Y depends on two or more explanatory variables, or regressors.

The simplest possible multiple regression model is three-variable regression, with one dependent variable and two explanatory variables. In this and the next chapter we shall study this model. Throughout, we are concerned with multiple linear regression models, that is, models linear in the parameters; they may or may not be linear in the variables.

7.1 THE THREE-VARIABLE MODEL: NOTATION AND ASSUMPTIONS

Generalizing the two-variable population regression function (PRF) (2.4.2), we may write the three-variable PRF as

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(7.1.1)}$$

where Y is the dependent variable, X_2 and X_3 the explanatory variables (or regressors), u the stochastic disturbance term, and i the ith observation; in case the data are time series, the subscript t will denote the tth observation.[1]

In Eq. (7.1.1) β_1 is the intercept term. As usual, it gives the mean or average effect on Y of all the variables excluded from the model, although its mechanical interpretation is the average value of Y when X_2 and X_3 are set equal to zero. The coefficients β_2 and β_3 are called the **partial regression coefficients,** and their meaning will be explained shortly.

We continue to operate within the framework of the classical linear regression model (CLRM) first introduced in Chapter 3. Specifically, we assume the following:

Zero mean value of u_i, or

$$E(u_i \mid X_{2i}, X_{3i}) = 0 \qquad \text{for each } i \tag{7.1.2}$$

No serial correlation, or

$$\operatorname{cov}(u_i, u_j) = 0 \qquad i \neq j \tag{7.1.3}$$

Homoscedasticity, or

$$\operatorname{var}(u_i) = \sigma^2 \tag{7.1.4}$$

Zero covariance between u_i and each X variable, or

$$\operatorname{cov}(u_i, X_{2i}) = \operatorname{cov}(u_i, X_{3i}) = 0 \tag{7.1.5}$$[2]

No specification bias, or

$$\text{The model is correctly specified} \tag{7.1.6}$$

No exact collinearity between the X variables, or

$$\text{No } \textbf{exact linear relationship} \text{ between } X_2 \text{ and } X_3 \tag{7.1.7}$$

In addition, as in Chapter 3, we assume that the multiple regression model is *linear in the parameters*, that the values of the regressors are fixed in repeated sampling, and that there is sufficient variability in the values of the regressors.

The rationale for assumptions (7.1.2) through (7.1.6) is the same as that discussed in Section 3.2. Assumption (7.1.7), that there be no exact linear relationship between X_2 and X_3, technically known as the assumption of

[1]For notational symmetry, Eq. (7.1.1) can also be written as

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

with the provision that $X_{1i} = 1$ for all i.

[2]This assumption is automatically fulfilled if X_2 and X_3 are nonstochastic and (7.1.2) holds.

no collinearity or **no multicollinearity** if more than one exact linear relationship is involved, is new and needs some explanation.

Informally, no collinearity means none of the regressors can be written as *exact* linear combinations of the remaining regressors in the model.

Formally, no collinearity means that there exists no set of numbers, λ_2 and λ_3, not both zero such that

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} = 0 \tag{7.1.8}$$

If such an exact linear relationship exists, then X_2 and X_3 are said to be **collinear** or linearly dependent. On the other hand, if (7.1.8) holds true only when $\lambda_2 = \lambda_3 = 0$, then X_2 and X_3 are said to be *linearly independent*.

Thus, if

$$X_{2i} = -4X_{3i} \qquad \text{or} \qquad X_{2i} + 4X_{3i} = 0 \tag{7.1.9}$$

the two variables are linearly dependent, and if both are included in a regression model, we will have perfect collinearity or an exact linear relationship between the two regressors.

Although we shall consider the problem of multicollinearity in depth in Chapter 10, intuitively the logic behind the assumption of no multicollinearity is not too difficult to grasp. Suppose that in (7.1.1) Y, X_2, and X_3 represent consumption expenditure, income, and wealth of the consumer, respectively. In postulating that consumption expenditure is linearly related to income and wealth, economic theory presumes that wealth and income may have some independent influence on consumption. If not, there is no sense in including both income and wealth variables in the model. In the extreme, if there is an exact linear relationship between income and wealth, we have only one independent variable, not two, and there is no way to assess the *separate* influence of income and wealth on consumption. To see this clearly, let $X_{3i} = 2X_{2i}$ in the consumption–income–wealth regression. Then the regression (7.1.1) becomes

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 (2X_{2i}) + u_i \\ &= \beta_1 + (\beta_2 + 2\beta_3) X_{2i} + u_i \\ &= \beta_1 + \alpha X_{2i} + u_i \end{aligned} \tag{7.1.10}$$

where $\alpha = (\beta_2 + 2\beta_3)$. That is, we in fact have a two-variable and not a three-variable regression. Moreover, if we run the regression (7.1.10) and obtain α, there is no way to estimate the separate influence of X_2 ($= \beta_2$) and X_3 ($= \beta_3$) on Y, for α gives the *combined influence* of X_2 and X_3 on Y.[3]

[3]Mathematically speaking, $\alpha = (\beta_2 + 2\beta_3)$ is one equation in two unknowns and there is no *unique* way of estimating β_2 and β_3 from the estimated α.

In short the assumption of no multicollinearity requires that in the PRF we include only those variables that are not exact linear functions of one or more variables in the model. Although we will discuss this topic more fully in Chapter 10, a couple of points may be noted here.

First, the assumption of no multicollinearity pertains to our theoretical (i.e., PRF) model. In practice, when we collect data for empirical analysis there is no guarantee that there will not be correlations among the regressors. As a matter of fact, in most applied work it is almost impossible to find two or more (economic) variables that may not be correlated to some extent, as we will show in our illustrative examples later in the chapter. What we require is that there be no exact relationships among the regressors, as in Eq. (7.1.9).

Second, keep in mind that we are talking only about perfect *linear* relationships between two or more variables. Multicollinearity does not rule out *nonlinear* relationships between variables. Suppose $X_{3i} = X_{2i}^2$. This does not violate the assumption of no perfect collinearity, as the relationship between the variables here is nonlinear.

7.2 INTERPRETATION OF MULTIPLE REGRESSION EQUATION

Given the assumptions of the classical regression model, it follows that, on taking the conditional expectation of Y on both sides of (7.1.1), we obtain

$$E(Y_i \mid X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i} \qquad \textbf{(7.2.1)}$$

In words, (7.2.1) gives the **conditional mean or expected value of *Y* conditional upon the given or fixed values of** X_2 and X_3. Therefore, as in the two-variable case, multiple regression analysis is regression analysis conditional upon the fixed values of the regressors, and what we obtain is the average or mean value of Y or the mean response of Y for the given values of the regressors.

7.3 THE MEANING OF PARTIAL REGRESSION COEFFICIENTS

As mentioned earlier, the regression coefficients β_2 and β_3 are known as **partial regression** or **partial slope coefficients.** The meaning of partial regression coefficient is as follows: β_2 measures the *change* in the mean value of Y, $E(Y)$, per unit change in X_2, holding the value of X_3 constant. Put differently, it gives the "direct" or the "net" effect of a unit change in X_2 on the mean value of Y, net of any effect that X_3 may have on mean Y. Likewise, β_3 measures the change in the mean value of Y per unit change in X_3, holding the value of X_2 constant.[4] That is, it gives the "direct" or "net" effect of a unit

[4]The calculus-minded reader will notice at once that β_2 and β_3 are the partial derivatives of $E(Y|X_2, X_3)$ with respect to X_2 and X_3.

change in X_3 on the mean value of Y, net of any effect that X_2 may have on mean Y.[5]

How do we actually go about holding the influence of a regressor constant? To explain this, let us revert to our child mortality example. Recall that in that example, $Y =$ child mortality (CM), $X_2 =$ per capita GNP (PGNP), and $X_3 =$ female literacy rate (FLR). Let us suppose we want to hold the influence of FLR constant. Since FLR may have some effect on CM as well as PGNP in any given concrete data, what we can do is to remove the (linear) influence of FLR from both CM and PGNP by running the regression of CM on FLR and that of PGNP on FLR separately and then looking at the residuals obtained from these regressions. Using the data given in Table 6.4, we obtain the following regressions:

$$\begin{aligned} \text{CM}_i &= 263.8635 \;-\; 2.3905\ \text{FLR}_i + \hat{u}_{1i} \\ \text{se} &= (12.2249) \quad (0.2133) \qquad\qquad r^2 = 0.6695 \end{aligned} \tag{7.3.1}$$

where $\hat{u}_{1i}$ represents the residual term of this regression.

$$\begin{aligned} \text{PGNP}_i &= -39.3033 \;+\; 28.1427\ \text{FLR}_i + \hat{u}_{2i} \\ \text{se} &= (734.9526) \quad (12.8211) \qquad\qquad r^2 = 0.0721 \end{aligned} \tag{7.3.2}$$

where $\hat{u}_{2i}$ represents the residual term of this regression.

Now

$$\hat{u}_{1i} = (\text{CM}_i - 263.8635 + 2.3905\ \text{FLR}_i) \tag{7.3.3}$$

represents that part of CM left after removing from it the (linear) influence of FLR. Likewise,

$$\hat{u}_{2i} = (\text{PGNP}_i + 39.3033 - 28.1427\ \text{FLR}_i) \tag{7.3.4}$$

represents that part of PGNP left after removing from it the (linear) influence of FLR.

Therefore, if we now regress $\hat{u}_{1i}$ on $\hat{u}_{2i}$, which are "purified" of the (linear) influence of FLR, wouldn't we obtain the net effect of PGNP on CM? That is indeed the case (see Appendix 7A, Section 7A.2). The regression results are as follows:

$$\begin{aligned} \hat{\hat{u}}_{1i} &= -0.0056\hat{u}_{2i} \\ \text{se} &= \quad (0.0019) \qquad r^2 = 0.1152 \end{aligned} \tag{7.3.5}$$

Note: This regression has no intercept term because the mean value of the OLS residuals $\hat{u}_{1i}$ and $\hat{u}_{2i}$ is zero (why?)

[5]Incidentally, the terms *holding constant, controlling for, allowing or accounting for the influence of, correcting the influence of*, and *sweeping out the influence of* are synonymous and will be used interchangeably in this text.

The slope coefficient of −0.0056 now gives the "true" or net effect of a unit change in PGNP on CM or the true slope of CM with respect to PGNP. That is, it gives the partial regression coefficient of CM with respect to PGNP, β_2.

Readers who want to get the partial regression coefficient of CM with respect to FLR can replicate the above procedure by first regressing CM on PGNP and getting the residuals from this regression ($\hat{u}_{1i}$), then regressing FLR on PGNP and obtaining the residuals from this regression ($\hat{u}_{2i}$), and then regressing $\hat{u}_{1i}$ on $\hat{u}_{2i}$. I am sure readers get the idea.

Do we have to go through this multistep procedure every time we want to find out the true partial regression coefficient? Fortunately, we do not have to do that, for the same job can be accomplished fairly quickly and routinely by the OLS procedure discussed in the next section. The multistep procedure just outlined is merely for pedagogic purposes to drive home the meaning of "partial" regression coefficient.

7.4 OLS AND ML ESTIMATION OF THE PARTIAL REGRESSION COEFFICIENTS

To estimate the parameters of the three-variable regression model (7.1.1), we first consider the method of ordinary least squares (OLS) introduced in Chapter 3 and then consider briefly the method of maximum likelihood (ML) discussed in Chapter 4.

OLS Estimators

To find the OLS estimators, let us first write the sample regression function (SRF) corresponding to the PRF of (7.1.1) as follows:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \qquad \textbf{(7.4.1)}$$

where $\hat{u}_i$ is the residual term, the sample counterpart of the stochastic disturbance term u_i.

As noted in Chapter 3, the OLS procedure consists in so choosing the values of the unknown parameters that the residual sum of squares (RSS) $\sum \hat{u}_i^2$ is as small as possible. Symbolically,

$$\min \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \qquad \textbf{(7.4.2)}$$

where the expression for the RSS is obtained by simple algebraic manipulations of (7.4.1).

The most straightforward procedure to obtain the estimators that will minimize (7.4.2) is to differentiate it with respect to the unknowns, set the resulting expressions to zero, and solve them simultaneously. As shown in Appendix 7A, Section 7A.1, this procedure gives the following *normal equations* [cf. Eqs. (3.1.4) and (3.1.5)]:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 \tag{7.4.3}$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} \tag{7.4.4}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 \tag{7.4.5}$$

From Eq. (7.4.3) we see at once that

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \tag{7.4.6}$$

which is the OLS estimator of the population intercept β_1.

Following the convention of letting the lowercase letters denote deviations from sample mean values, one can derive the following formulas from the normal equations (7.4.3) to (7.4.5):

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \tag{7.4.7}$$[6]

$$\hat{\beta}_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \tag{7.4.8}$$

which give the OLS estimators of the population partial regression coefficients β_2 and β_3, respectively.

In passing, note the following: (1) Equations (7.4.7) and (7.4.8) are symmetrical in nature because one can be obtained from the other by interchanging the roles of X_2 and X_3; (2) the denominators of these two equations are identical; and (3) the three-variable case is a natural extension of the two-variable case.

Variances and Standard Errors of OLS Estimators

Having obtained the OLS estimators of the partial regression coefficients, we can derive the variances and standard errors of these estimators in the manner indicated in Appendix 3A.3. As in the two-variable case, we need the standard errors for two main purposes: to establish confidence intervals and

[6]This estimator is equal to that of (7.3.5), as shown in App. 7A, Sec. 7A.2.

to test statistical hypotheses. The relevant formulas are as follows:[7]

$$\operatorname{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2\bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i}x_{3i}\right)^2}\right] \cdot \sigma^2 \quad \textbf{(7.4.9)}$$

$$\operatorname{se}(\hat{\beta}_1) = +\sqrt{\operatorname{var}(\hat{\beta}_1)} \quad \textbf{(7.4.10)}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2 \quad \textbf{(7.4.11)}$$

or, equivalently,

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2\left(1 - r_{23}^2\right)} \quad \textbf{(7.4.12)}$$

where r_{23} is the sample coefficient of correlation between X_2 and X_3 as defined in Chapter 3.[8]

$$\operatorname{se}(\hat{\beta}_2) = +\sqrt{\operatorname{var}(\hat{\beta}_2)} \quad \textbf{(7.4.13)}$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2 \quad \textbf{(7.4.14)}$$

or, equivalently,

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2\left(1 - r_{23}^2\right)} \quad \textbf{(7.4.15)}$$

$$\operatorname{se}(\hat{\beta}_3) = +\sqrt{\operatorname{var}(\hat{\beta}_3)} \quad \textbf{(7.4.16)}$$

$$\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{\left(1 - r_{23}^2\right)\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}} \quad \textbf{(7.4.17)}$$

In all these formulas σ^2 is the (homoscedastic) variance of the population disturbances u_i.

Following the argument of Appendix 3A, Section 3A.5, the reader can verify that an unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - 3} \quad \textbf{(7.4.18)}$$

[7]The derivations of these formulas are easier using matrix notation. Advanced readers may refer to **App. C.**

[8]Using the definition of r given in Chap. 3, we have

$$r_{23}^2 = \frac{\left(\sum x_{2i}x_{3i}\right)^2}{\sum x_{2t}^2 \sum x_{3t}^2}$$

Note the similarity between this estimator of σ^2 and its two-variable counterpart [$\hat{\sigma}^2 = (\sum \hat{u}_i^2)/(n-2)$]. The degrees of freedom are now $(n-3)$ because in estimating $\sum \hat{u}_i^2$ we must first estimate β_1, β_2, and β_3, which consume 3 df. (The argument is quite general. Thus, in the four-variable case the df will be $n-4$.)

The estimator $\hat{\sigma}^2$ can be computed from (7.4.18) once the residuals are available, but it can also be obtained more readily by using the following relation (for proof, see Appendix 7A, Section 7A.3):

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

which is the three-variable counterpart of the relation given in (3.3.6).

Properties of OLS Estimators

The properties of OLS estimators of the multiple regression model parallel those of the two-variable model. Specifically:

1. The three-variable regression line (surface) passes through the means $\bar{Y}$, $\bar{X}_2$, and $\bar{X}_3$, which is evident from (7.4.3) [cf. Eq. (3.1.7) of the two-variable model]. This property holds generally. Thus in the k-variable linear regression model [a regressand and $(k-1)$ regressors]

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{7.4.20}$$

we have

$$\hat{\beta}_1 = \bar{Y} - \beta_2 \bar{X}_2 - \beta_3 \hat{X}_3 - \cdots - \beta_k \bar{X}_k \tag{7.4.21}$$

2. The mean value of the estimated Y_i $(= \hat{Y}_i)$ is equal to the mean value of the actual Y_i, which is easy to prove:

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3) + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \qquad \text{(Why?)} \\ &= \bar{Y} + \hat{\beta}_2 (X_{2i} - \bar{X}_2) + \hat{\beta}_3 (X_{3i} - \bar{X}_3) \\ &= \bar{Y} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \end{aligned} \tag{7.4.22}$$

where as usual small letters indicate values of the variables as deviations from their respective means.

Summing both sides of (7.4.22) over the sample values and dividing through by the sample size n gives $\bar{\hat{Y}} = \bar{Y}$. (*Note:* $\sum x_{2i} = \sum x_{3i} = 0$. Why?) Notice that by virtue of (7.4.22) we can write

$$\hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \tag{7.4.23}$$

where $\hat{y}_i = (\hat{Y}_i - \bar{Y})$.

Therefore, the SRF (7.4.1) can be expressed in the *deviation form* as

$$y_i = \hat{y}_i + \hat{u}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \tag{7.4.24}$$

3. $\sum \hat{u}_i = \bar{\hat{u}} = 0$, which can be verified from (7.4.24). [*Hint:* Sum both sides of (7.4.24) over the sample values.]

4. The residuals $\hat{u}_i$ are uncorrelated with X_{2i} and X_{3i}, that is, $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$ (see Appendix 7A.1 for proof).

5. The residuals $\hat{u}_i$ are uncorrelated with $\hat{Y}_i$; that is, $\sum \hat{u}_i \hat{Y}_i = 0$. Why? [*Hint:* Multiply (7.4.23) on both sides by $\hat{u}_i$ and sum over the sample values.]

6. From (7.4.12) and (7.4.15) it is evident that as r_{23}, the correlation coefficient between X_2 and X_3, increases toward 1, the variances of $\hat{\beta}_2$ and $\hat{\beta}_3$ increase for given values of σ^2 and $\sum x_{2i}^2$ or $\sum x_{3i}^2$. In the limit, when $r_{23} = 1$ (i.e., perfect collinearity), these variances become infinite. The implications of this will be explored fully in Chapter 10, but intuitively the reader can see that as r_{23} increases it is going to be increasingly difficult to know what the true values of β_2 and β_3 are. [More on this in the next chapter, but refer to Eq. (7.1.10).]

7. It is also clear from (7.4.12) and (7.4.15) that for given values of r_{23} and $\sum x_{2i}^2$ or $\sum x_{3i}^2$, the variances of the OLS estimators are directly proportional to σ^2; that is, they increase as σ^2 increases. Similarly, for given values of σ^2 and r_{23}, the variance of $\hat{\beta}_2$ is inversely proportional to $\sum x_{2i}^2$; that is, the greater the variation in the sample values of X_2, the smaller the variance of $\hat{\beta}_2$ and therefore β_2 can be estimated more precisely. A similar statement can be made about the variance of $\hat{\beta}_3$.

8. Given the assumptions of the classical linear regression model, which are spelled out in Section 7.1, one can prove that the OLS estimators of the partial regression coefficients not only are linear and unbiased but also have minimum variance in the class of all linear unbiased estimators. In short, *they are BLUE:* Put differently, they satisfy the Gauss-Markov theorem. (The proof parallels the two-variable case proved in Appendix 3A, Section 3A.6 and will be presented more compactly using matrix notation in **Appendix C.**)

Maximum Likelihood Estimators

We noted in Chapter 4 that under the assumption that u_i, the population disturbances, are normally distributed with zero mean and constant variance σ^2, the maximum likelihood (ML) estimators and the OLS estimators of the regression coefficients of the two-variable model are identical. This equality extends to models containing any number of variables. (For proof, see Appendix 7A, Section 7A.4.) However, this is not true of the estimator of σ^2. It can be shown that the ML estimator of σ^2 is $\sum \hat{u}_i^2/n$ regardless of the number of variables in the model, whereas the OLS estimator of σ^2 is $\sum \hat{u}_i^2/(n-2)$ in the two-variable case, $\sum \hat{u}_i^2/(n-3)$ in the three-variable case, and $\sum \hat{u}_i^2/(n-k)$ in the case of the k-variable model (7.4.20). In short, the OLS estimator of σ^2 takes into account the number of degrees of freedom, whereas the ML estimator does not. Of course, if n is very large, the ML and OLS estimators of σ^2 will tend to be close to each other. (Why?)

7.5 THE MULTIPLE COEFFICIENT OF DETERMINATION R^2 AND THE MULTIPLE COEFFICIENT OF CORRELATION R

In the two-variable case we saw that r^2 as defined in (3.5.5) measures the goodness of fit of the regression equation; that is, it gives the proportion or percentage of the total variation in the dependent variable Y explained by the (single) explanatory variable X. This notation of r^2 can be easily extended to regression models containing more than two variables. Thus, in the three-variable model we would like to know the proportion of the variation in Y explained by the variables X_2 and X_3 jointly. The quantity that gives this information is known as the **multiple coefficient of determination** and is denoted by R^2; conceptually it is akin to r^2.

To derive R^2, we may follow the derivation of r^2 given in Section 3.5. Recall that

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned} \tag{7.5.1}$$

where $\hat{Y}_i$ is the estimated value of Y_i from the fitted regression line and is an estimator of true $E(Y_i \mid X_{2i}, X_{3i})$. Upon shifting to lowercase letters to indicate deviations from the mean values, Eq. (7.5.1) may be written as

$$\begin{aligned} y_i &= \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \\ &= \hat{y}_i + \hat{u}_i \end{aligned} \tag{7.5.2}$$

Squaring (7.5.2) on both sides and summing over the sample values, we obtain

$$\begin{aligned} \sum y_i^2 &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2 \sum \hat{y}_i \hat{u}_i \\ &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \qquad \text{(Why?)} \end{aligned} \tag{7.5.3}$$

Verbally, Eq. (7.5.3) states that the total sum of squares (TSS) equals the explained sum of squares (ESS) + the residual sum of squares (RSS). Now substituting for $\sum \hat{u}_i^2$ from (7.4.19), we obtain

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

which, on rearranging, gives

$$\text{ESS} = \sum \hat{y}_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} \tag{7.5.4}$$

Now, by definition

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2} \qquad \textbf{(7.5.5)}^9$$

[cf. (7.5.5) with (3.5.6)].

Since the quantities entering (7.5.5) are generally computed routinely, R^2 can be computed easily. Note that R^2, like r^2, lies between 0 and 1. If it is 1, the fitted regression line explains 100 percent of the variation in Y. On the other hand, if it is 0, the model does not explain any of the variation in Y. Typically, however, R^2 lies between these extreme values. The fit of the model is said to be "better" the closer is R^2 to 1.

Recall that in the two-variable case we defined the quantity r as the coefficient of correlation and indicated that it measures the degree of (linear) association between two variables. The three-or-more-variable analogue of r is the coefficient of **multiple correlation,** denoted by R, and it is a measure of the degree of association between Y and all the explanatory variables jointly. Although r can be positive or negative, R is always taken to be positive. In practice, however, R is of little importance. The more meaningful quantity is R^2.

Before proceeding further, let us note the following relationship between R^2 and the variance of a partial regression coefficient in the k-variable multiple regression model given in (7.4.20):

$$\text{var}\,(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1 - R_j^2}\right) \qquad \textbf{(7.5.6)}$$

where $\hat{\beta}_j$ is the partial regression coefficient of regressor X_j and R_j^2 is the R^2 in the regression of X_j on the remaining $(k - 2)$ regressors. [*Note:* There are $(k - 1)$ regressors in the k-variable regression model.] Although the utility of Eq. (7.5.6) will become apparent in Chapter 10 on multicollinearity, observe that this equation is simply an extension of the formula given in (7.4.12) or (7.4.15) for the three-variable regression model, one regressand and two regressors.

7.6 EXAMPLE 7.1: CHILD MORTALITY IN RELATION TO PER CAPITA GNP AND FEMALE LITERACY RATE

In Chapter 6 we considered the behavior of child mortality (CM) in relation to per capita GNP (PGNP). There we found that PGNP has a negative impact on CM, as one would expect. Now let us bring in female literacy as measured

[9]Note that R^2 can also be computed as follows:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{(n-3)\hat{\sigma}^2}{(n-1)S_y^2}$$

by the female literacy rate (FLR). A priori, we expect that FLR too will have a negative impact on CM. Now when we introduce both the variables in our model, we need to net out the influence of each of the regressors. That is, we need to estimate the (partial) regression coefficients of each regressor. Thus our model is:

$$\text{CM}_i = \beta_1 + \beta_2 \text{PGNP}_i + \beta_3 \text{FLR}_i + u_i \qquad \textbf{(7.6.1)}$$

The necessary data are given in Table 6.4. Keep in mind that CM is the number of deaths of children under five per 1000 live births, PGNP is per capita GNP in 1980, and FLR is measured in percent. Our sample consists of 64 countries.

Using the Eviews3 statistical package, we obtained the following results:

$$\widehat{\text{CM}}_i = 263.6416 - 0.0056\ \text{PGNP}_i - 2.2316\ \text{FLR}_i$$
$$\text{se} = (11.5932) \quad (0.0019) \qquad (0.2099) \qquad R^2 = 0.7077 \qquad \textbf{(7.6.2)}$$
$$\bar{R}^2 = 0.6981^*$$

where figures in parentheses are the estimated standard errors. Before we interpret this regression, observe the partial slope coefficient of PGNP, namely, −0.0056. Is it not precisely the same as that obtained from the three-step procedure discussed in the previous section [see Eq. (7.3.5)]? But should that surprise you? Not only that, but the two standard errors are precisely the same, which is again unsurprising. But we did so without the three-step cumbersome procedure.

Let us now interpret these regression coefficients: −0.0056 is the partial regression coefficient of PGNP and tells us that with the influence of FLR held constant, as PGNP increases, say, by a dollar, on average, child mortality goes down by 0.0056 units. To make it more economically interpretable, if the per capita GNP goes up by a thousand dollars, on average, the number of deaths of children under age 5 goes down by about 5.6 per thousand live births. The coefficient −2.2316 tells us that holding the influence of PGNP constant, on average, the number of deaths of children under 5 goes down by about 2.23 per thousand live births as the female literacy rate increases by one percentage point. The intercept value of about 263, mechanically interpreted, means that if the values of PGNP and FLR rate were fixed at zero, the mean child mortality would be about 263 deaths per thousand live births. Of course, such an interpretation should be taken with a grain of salt. All one could infer is that if the two regressors were fixed at zero, child mortality will be quite high, which makes practical sense. The R^2 value of about 0.71 means that about 71 percent of the variation in child mortality is explained by PGNP and FLR, a fairly high value considering that the maximum value of R^2 can at most be 1. All told, the regression results make sense.

*On this, see Sec. 7.8.

What about the statistical significance of the estimated coefficients? We will take this topic up in Chapter 8. As we will see there, in many ways this chapter will be an extension of Chapter 5, which dealt with the two-variable model. As we will also show, there are some important differences in statistical inference (i.e., hypothesis testing) between the two-variable and multivariable regression models.

Regression on Standardized Variables

In the preceding chapter we introduced the topic of regression on standardized variables and stated that the analysis can be extended to multivariable regressions. Recall that a variable is said to be standardized or in standard deviation units if it is expressed in terms of deviation from its mean and divided by its standard deviation.

For our child mortality example, the results are as follows:

$$\begin{aligned} \widehat{\text{CM}}^{*} &= -\ 0.2026\ \text{PGNP}_i^{*} - 0.7639\ \text{FLR}_i^{*} \\ \text{se} &= \quad (0.0713) \qquad\quad (0.0713) \qquad\quad r^2 = 0.7077 \end{aligned} \tag{7.6.3}$$

Note: The starred variables are standardized variables. Also note that there is no intercept in the model for reasons already discussed in the previous chapter.

As you can see from this regression, with FLR held constant, a standard deviation increase in PGNP leads, on average, to a 0.2026 standard deviation decrease in CM. Similarly, holding PGNP constant, a standard deviation increase in FLR, on average, leads to a 0.7639 standard deviation decrease in CM. Relatively speaking, female literacy has more impact on child mortality than per capita GNP. Here you will see the advantage of using standardized variables, for standardization puts all variables on equal footing because all standardized variables have zero means and unit variances.

7.7 SIMPLE REGRESSION IN THE CONTEXT OF MULTIPLE REGRESSION: INTRODUCTION TO SPECIFICATION BIAS

Recall that assumption (7.1.6) of the classical linear regression model states that the regression model used in the analysis is "correctly" specified; that is, there is no **specification bias or specification error** (see Chapter 3 for some introductory remarks). Although the topic of specification error will be discussed more fully in Chapter 13, the illustrative example given in the preceding section provides a splendid opportunity not only to drive home the importance of assumption (7.1.6) but also to shed additional light on the meaning of partial regression coefficient and provide a somewhat informal introduction to the topic of specification bias.

Assume that (7.6.1) is the "true" model explaining the behavior of child mortality in relation to per capita GNP and female literacy rate (FLR). But suppose we disregard FLR and estimate the following simple regression:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i} \tag{7.7.1}$$

where $Y = \text{CM}$ and $X_2 = \text{PGNP}$.

Since (7.6.1) is the true model, estimating (7.7.1) would constitute a specification error; the error here consists in *omitting* the variable X_3, the female literacy rate. Notice that we are using different parameter symbols (the alphas) in (7.7.1) to distinguish them from the true parameters (the betas) given in (7.6.1).

Now will α_2 provide an unbiased estimate of the true impact of PGNP, which is given by β_2 in model (7.6.1)? In other words, will $E(\hat{\alpha}_2) = \beta_2$, where $\hat{\alpha}_2$ is the estimated value of α_2? In other words, will the coefficient of PGNP in (7.7.1) provide an unbiased estimate of the true impact of PGNP on CM, knowing that we have omitted the variable X_3 (FLR) from the model? As you would suspect, *in general* $\hat{\alpha}_2$ will not be an unbiased estimator of the true β_2. To give a glimpse of the bias, let us run the regression (7.7.1), which gave the following results.

$$\begin{aligned} \widehat{\text{CM}}_i &= 157.4244 \;-\; 0.0114\ \text{PGNP}_i \\ \text{se} &= \quad (9.8455) \quad (0.0032) \qquad\qquad r^2 = 0.1662 \end{aligned} \tag{7.7.2}$$

Observe several things about this regression compared to the "true" multiple regression (7.6.1):

1. In absolute terms (i.e., disregarding the sign), the PGNP coefficient has increased from 0.0056 to 0.0114, almost a two-fold increase.

2. The standard errors are different.

3. The intercept values are different.

4. The r^2 values are dramatically different, although it is generally the case that, as the number of regressors in the model increases, the r^2 value increases.

Now suppose that you regress child mortality on female literacy rate, disregarding the influence of PGNP. You will obtain the following results:

$$\begin{aligned} \widehat{\text{CM}}_i &= 263.8635 \;-\; 2.3905\ \text{FLR}_i \\ \text{se} &= (21.2249) \quad (0.2133) \qquad\qquad r^2 = 0.6696 \end{aligned} \tag{7.7.3}$$

Again if you compare the results of this (misspecified) regression with the "true" multiple regression, you will see that the results are different, although the difference here is not as noticeable as in the case of regression (7.7.2).

The important point to note is that serious consequences can ensue if you misfit a model. We will look into this topic more thoroughly in Chapter 13, on specification errors.

7.8 R^2 AND THE ADJUSTED R^2

An important property of R^2 is that it is a nondecreasing function of the number of explanatory variables or regressors present in the model; as the number of regressors increases, R^2 almost invariably increases and never decreases. Stated differently, an additional X variable will not decrease R^2. Compare, for instance, regression (7.7.2) or (7.7.3) with (7.6.2). To see this, recall the definition of the coefficient of determination:

$$\begin{aligned} R^2 &= \frac{\text{ESS}}{\text{TSS}} \\ &= 1 - \frac{\text{RSS}}{\text{TSS}} \\ &= 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \end{aligned} \tag{7.8.1}$$

Now $\sum y_i^2$ is independent of the number of X variables in the model because it is simply $\sum(Y_i - \bar{Y})^2$. The RSS, $\sum \hat{u}_i^2$, however, depends on the number of regressors present in the model. Intuitively, it is clear that as the number of X variables increases, $\sum \hat{u}_i^2$ is likely to decrease (at least it will not increase); hence R^2 as defined in (7.8.1) will increase. In view of this, in comparing two regression models with the *same dependent variable* but differing number of X variables, one should be very wary of choosing the model with the highest R^2.

To compare two R^2 terms, one must take into account the number of X variables present in the model. This can be done readily if we consider an alternative coefficient of determination, which is as follows:

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 / (n-k)}{\sum y_i^2 / (n-1)} \tag{7.8.2}$$

where $k =$ the number of parameters in the model *including the intercept term*. (In the three-variable regression, $k = 3$. Why?) The R^2 thus defined is known as the **adjusted R^2,** denoted by $\bar{R}^2$. The term *adjusted* means adjusted for the df associated with the sums of squares entering into (7.8.1): $\sum \hat{u}_i^2$ has $n - k$ df in a model involving k parameters, which include

the intercept term, and $\sum y_i^2$ has $n - 1$ df. (Why?) For the three-variable case, we know that $\sum \hat{u}_i^2$ has $n - 3$ df.

Equation (7.8.2) can also be written as

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{S_Y^2} \tag{7.8.3}$$

where $\hat{\sigma}^2$ is the residual variance, an unbiased estimator of true σ^2, and S_Y^2 is the sample variance of Y.

It is easy to see that $\bar{R}^2$ and R^2 are related because, substituting (7.8.1) into (7.8.2), we obtain

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{7.8.4}$$

It is immediately apparent from Eq. (7.8.4) that (1) for $k > 1$, $\bar{R}^2 < R^2$ which implies that as the number of X variables increases, the adjusted R^2 increases less than the unadjusted R^2; and (2) $\bar{R}^2$ can be negative, although R^2 is necessarily nonnegative.[10] In case $\bar{R}^2$ turns out to be negative in an application, its value is taken as zero.

Which R^2 should one use in practice? As Theil notes:

> . . . it is good practice to use $\bar{R}^2$ rather than R^2 because R^2 tends to give an overly optimistic picture of the fit of the regression, particularly when the number of explanatory variables is not very small compared with the number of observations.[11]

But Theil's view is not uniformly shared, for he has offered no general theoretical justification for the "superiority" of $\bar{R}^2$. For example, Goldberger argues that the following R^2, call it **modified** R^2, will do just as well[12]:

$$\text{Modified } R^2 = (1 - k/n)R^2 \tag{7.8.5}$$

His advice is to report R^2, n, and k and let the reader decide how to adjust R^2 by allowing for n and k.

Despite this advice, it is the adjusted R^2, as given in (7.8.4), that is reported by most statistical packages along with the conventional R^2. The reader is well advised to treat $\bar{R}^2$ as just another summary statistic.

---

[10]Note, however, that if $R^2 = 1$, $\bar{R}^2 = R^2 = 1$. When $R^2 = 0$, $\bar{R}^2 = (1 - k)/(n - k)$, in which case $\bar{R}^2$ can be negative if $k > 1$.

[11]Henri Theil, *Introduction to Econometrics,* Prentice Hall, Englewood Cliffs, N.J., 1978, p. 135.

[12]Arthur S. Goldberger, *A Course in Econometrics,* Harvard University Press, Cambridge, Mass., 1991, p. 178. For a more critical view of R^2, see S. Cameron, "Why is the *R* Squared Adjusted Reported?", *Journal of Quantitative Economics,* vol. 9, no. 1, January 1993, pp. 183–186. He argues that "It [R^2] is NOT a test statistic and there seems to be no clear intuitive justification for its use as a descriptive statistic. Finally, we should be clear that it is not an effective tool for the prevention of data mining" (p. 186).

Incidentally, for the child mortality regression (7.6.2), the reader should verify that $\bar{R}^2$ is 0.6981, keeping in mind that in this example $(n-1)=63$ and $(n-k)=60$. As expected, $\bar{R}^2$ of 0.6981 is less than R^2 of 0.7077.

Besides R^2 and adjusted R^2 as goodness of fit measures, other criteria are often used to judge the adequacy of a regression model. Two of these are **Akaike's Information criterion** and **Amemiya's Prediction criteria,** which are used to select between competing models. We will discuss these criteria when we consider the problem of model selection in greater detail in a later chapter (see Chapter 13).

Comparing Two R^2 Values

It is crucial to note that in comparing two models on the basis of the coefficient of determination, whether adjusted or not, *the sample size n and the dependent variable must be the same*; the explanatory variables may take any form. Thus for the models

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(7.8.6)}$$

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i \qquad \textbf{(7.8.7)}$$

the computed R^2 terms cannot be compared. The reason is as follows: By definition, R^2 measures the proportion of the variation in the dependent variable accounted for by the explanatory variable(s). Therefore, in (7.8.6) R^2 measures the proportion of the *variation in* $\ln Y$ explained by X_2 and X_3, whereas in (7.8.7) it measures the proportion of the *variation in* Y, and the two are not the same thing: As noted in Chapter 6, a change in $\ln Y$ gives a relative or proportional change in Y, whereas a change in Y gives an absolute change. Therefore, $\text{var}\,\hat{Y}_i/\text{var}\,Y_i$ is not equal to $\text{var}\,(\widehat{\ln Y_i})/\text{var}\,(\ln Y_i)$; that is, the two coefficients of determination are not the same.[13]

How then does one compare the R^2's of two models when the regressand is not in the same form? To answer this question, let us first consider a numerical example.

[13]From the definition of R^2, we know that

$$1 - R^2 = \frac{\text{RSS}}{\text{TSS}} = \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

for the linear model and

$$1 - R^2 = \frac{\sum \hat{u}_i^2}{\sum (\ln Y_i - \overline{\ln Y})^2}$$

for the log model. Since the denominators on the right-hand sides of these expressions are different, we cannot compare the two R^2 terms directly.

As shown in Example 7.2, for the linear specification, the RSS = 0.1491 (the residual sum of squares of coffee consumption), and for the log–linear specification, the RSS = 0.0226 (the residual sum of squares of log of coffee consumption). These residuals are of different orders of magnitude and hence are not directly comparable.

EXAMPLE 7.2

COFFEE CONSUMPTION IN THE UNITED STATES, 1970–1980

Consider the data in Table 7.1. The data pertain to consumption of cups of coffee per day (Y) and real retail price of coffee (X) in the United States for years 1970–1980. Applying OLS to the data, we obtain the following regression results:

$$\hat{Y}_t = 2.6911 - 0.4795X_t \qquad \textbf{(7.8.8)}$$
$$\text{se} = (0.1216) \quad (0.1140) \qquad \text{RSS} = 0.1491;\ r^2 = 0.6628$$

The results make economic sense: As the price of coffee increases, on average, coffee consumption goes down by about half a cup per day. The r^2 value of about 0.66 means that the price of coffee explains about 66 percent of the variation in coffee consumption. The reader can readily verify that the slope coefficient is statistically significant.

From the same data, the following double log, or constant elasticity, model can be estimated:

$$\widehat{\ln Y}_t = 0.7774 - 0.2530 \ln X_t \qquad \textbf{(7.8.9)}$$
$$\text{se} = (0.0152) \quad (0.0494) \qquad \text{RSS} = 0.0226;\ r^2 = 0.7448$$

Since this is a double log model, the slope coefficient gives a direct estimate of the price elasticity coefficient. In the present instance, it tells us that if the price of coffee per pound goes up by 1 percent, on average, per day coffee consumption goes down by about 0.25 percent. Remember that in the linear model (7.8.8) the slope coefficient only gives the rate of change of coffee consumption with respect to price. (How will you estimate the price elasticity for the

TABLE 7.1
U.S. COFFEE CONSUMPTION (Y) IN RELATION TO AVERAGE REAL RETAIL PRICE (X),* 1970–1980

Year	Y, Cups per person per day	X, $ per lb
1970	2.57	0.77
1971	2.50	0.74
1972	2.35	0.72
1973	2.30	0.73
1974	2.25	0.76
1975	2.20	0.75
1976	2.11	1.08
1977	1.94	1.81
1978	1.97	1.39
1979	2.06	1.20
1980	2.02	1.17

**Note:* The nominal price was divided by the Consumer Price Index (CPI) for food and beverages, 1967 = 100.

Source: The data for Y are from *Summary of National Coffee Drinking Study*, Data Group, Elkins Park, Penn., 1981; and the data on nominal X (i.e., X in current prices) are from *Nielsen Food Index,* A. C. Nielsen, New York, 1981.

I am indebted to Scott E. Sandberg for collecting the data.

(*Continued*)

EXAMPLE 7.2 (*Continued*)

linear model?) The r^2 value of about 0.74 means that about 74 percent of the variation in the log of coffee demand is explained by the variation in the log of coffee price.

Since the r^2 value of the linear model of 0.6628 is smaller than the r^2 value of 0.7448 of the log–linear model, you might be tempted to choose the latter model because of its high r^2 value. But for reasons already noted, we cannot do so. But if you do want to compare the two r^2 values, you may proceed as follows:

1. Obtain $\widehat{\ln Y_t}$ from (7.8.9) for each observation; that is, obtain the estimated log value of each observation from this model. Take the antilog of these values and then compute r^2 between these antilog values and actual Y_t in the manner indicated by Eq. (3.5.14). This r^2 value is comparable to the r^2 value of the linear model (7.8.8).

2. *Alternatively,* assuming all Y values are positive, take logarithms of the Y values, ln Y. Obtain the estimated Y values, $\hat{Y}_t$, from the linear model (7.8.8), take the logarithms of these estimated Y values (i.e., ln $\hat{Y}_t$) and compute the r^2 between (ln Y_t) and (ln $\hat{Y}_t$) in the manner indicated in Eq. (3.5.14). This r^2 value is comparable to the r^2 value obtained from (7.8.9).

For our coffee example, we present the necessary raw data to compute the comparable r^2's in Table 7.2. To compare the r^2 value of the linear model (7.8.8) with that of (7.8.9), we first obtain log of ($\hat{Y}_t$) [given in column (6) of Table 7.2], then we obtain the log of actual Y values [given in column (5) of the table], and then compute r^2 between these two sets of values using Eq. (3.5.14). The result is an r^2 value of 0.7318, which is now comparable with the r^2 value of the log–linear model of 0.7448. Now the difference between the two r^2 values is very small.

On the other hand, if we want to compare the r^2 value of the log–linear model with the linear model, we obtain $\widehat{\ln Y_t}$ for each observation from (7.8.9) [given in column (3) of the table], obtain their antilog values [given in column (4) of the table], and finally compute r^2 between these antilog values and the actual Y values, using formula (3.5.14). This will give an r^2 value of 0.7187, which is slightly higher than that obtained from the linear model (7.8.8), namely, 0.6628.

Using either method, it seems that the log–linear model gives a slightly better fit.

TABLE 7.2 RAW DATA FOR COMPARING TWO R^2 VALUES

Year	Y_t (1)	$\hat{Y}_t$ (2)	$\widehat{\ln Y_t}$ (3)	Antilog of $\widehat{\ln Y_t}$ (4)	ln Y_t (5)	ln ($\hat{Y}_t$) (6)
1970	2.57	2.321887	0.843555	2.324616	0.943906	0.842380
1971	2.50	2.336272	0.853611	2.348111	0.916291	0.848557
1972	2.35	2.345863	0.860544	2.364447	0.854415	0.852653
1973	2.30	2.341068	0.857054	2.356209	0.832909	0.850607
1974	2.25	2.326682	0.846863	2.332318	0.810930	0.844443
1975	2.20	2.331477	0.850214	2.340149	0.788457	0.846502
1976	2.11	2.173233	0.757943	2.133882	0.746688	0.776216
1977	1.94	1.823176	0.627279	1.872508	0.662688	0.600580
1978	1.97	2.024579	0.694089	2.001884	0.678034	0.705362
1979	2.06	2.115689	0.731282	2.077742	0.722706	0.749381
1980	2.02	2.130075	0.737688	2.091096	0.703098	0.756157

Notes: Column (1): Actual Y values from Table 7.1
Column (2): Estimated Y values from the linear model (7.8.8)
Column (3): Estimated log Y values from the double-log model (7.8.9)
Column (4): Antilog of values in column (3)
Column (5): Log values of Y in column (1)
Column (6): Log values of $\hat{Y}_t$ in column (2)

Allocating R^2 among Regressors

Let us return to our child mortality example. We saw in (7.6.2) that the two regressors PGNP and FLR explain 0.7077 or 70.77 percent of the variation in child mortality. But now consider the regression (7.7.2) where we dropped the FLR variable and as a result the r^2 value dropped to 0.1662. Does that mean the difference in the r^2 value of 0.5415 (0.7077 − 0.1662) is attributable to the dropped variable FLR? On the other hand, if you consider regression (7.7.3), where we dropped the PGNP variable, the r^2 value drops to 0.6696. Does that mean the difference in the r^2 value of 0.0381 (0.7077 − 0.6696) is due to the omitted variable PGNP?

The question then is: Can we allocate the multiple R^2 of 0.7077 between the two regressors, PGNP and FLR, in this manner? Unfortunately, we cannot do so, for the allocation depends on the order in which the regressors are introduced, as we just illustrated. Part of the problem here is that the two regressors are correlated, the correlation coefficient between the two being 0.2685 (verify it from the data given in Table 6.4). In most applied work with several regressors, correlation among them is a common problem. Of course, the problem will be very serious if there is perfect collinearity among the regressors.

The best practical advice is that there is little point in trying to allocate the R^2 value to its constituent regressors.

The "Game" of Maximizing $\bar{R}^2$

In concluding this section, a warning is in order: Sometimes researchers play the game of maximizing $\bar{R}^2$, that is, choosing the model that gives the highest $\bar{R}^2$. But this may be dangerous, for in regression analysis our objective is not to obtain a high $\bar{R}^2$ per se but rather to obtain dependable estimates of the true population regression coefficients and draw statistical inferences about them. In empirical analysis it is not unusual to obtain a very high $\bar{R}^2$ but find that some of the regression coefficients either are statistically insignificant or have signs that are contrary to a priori expectations. Therefore, the researcher should be more concerned about the logical or theoretical relevance of the explanatory variables to the dependent variable and their statistical significance. If in this process we obtain a high $\bar{R}^2$, well and good; on the other hand, if $\bar{R}^2$ is low, it does not mean the model is necessarily bad.[14]

[14]Some authors would like to deemphasize the use of R^2 as a measure of goodness of fit as well as its use for comparing two or more R^2 values. See Christopher H. Achen, *Interpreting and Using Regression,* Sage Publications, Beverly Hills, Calif., 1982, pp. 58–67, and C. Granger and P. Newbold, "R^2 and the Transformation of Regression Variables," *Journal of Econometrics,* vol. 4, 1976, pp. 205–210. Incidentally, the practice of choosing a model on the basis of highest R^2, a kind of data mining, introduces what is known as **pretest bias,** which might destroy some of the properties of OLS estimators of the classical linear regression model. On this topic, the reader may want to consult George G. Judge, Carter R. Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee, *Introduction to the Theory and Practice of Econometrics,* John Wiley, New York, 1982, Chap. 21.

As a matter of fact, Goldberger is very critical about the role of R^2. He has said:

> From our perspective, R^2 has a very modest role in regression analysis, being a measure of the goodness of fit of a sample LS [least-squares] linear regression in a body of data. Nothing in the CR [CLRM] model requires that R^2 be high. Hence a high R^2 is not evidence in favor of the model and a low R^2 is not evidence against it.
>
> In fact the most important thing about R^2 is that it is not important in the CR model. The CR model is concerned with parameters in a population, not with goodness of fit in the sample. . . . If one insists on a measure of predictive success (or rather failure), then σ^2 might suffice: after all, the parameter σ^2 is the expected squared forecast error that would result if the population CEF [PRF] were used as the predictor. Alternatively, the squared standard error of forecast . . . at relevant values of $\boldsymbol{x}$ [regressors] may be informative.[15]

7.9 EXAMPLE 7.3: THE COBB–DOUGLAS PRODUCTION FUNCTION: MORE ON FUNCTIONAL FORM

In Section 6.4 we showed how with appropriate transformations we can convert nonlinear relationships into linear ones so that we can work within the framework of the classical linear regression model. The various transformations discussed there in the context of the two-variable case can be easily extended to multiple regression models. We demonstrate transformations in this section by taking up the multivariable extension of the two-variable log–linear model; others can be found in the exercises and in the illustrative examples discussed throughout the rest of this book. The specific example we discuss is the celebrated **Cobb–Douglas production function** of production theory.

The Cobb–Douglas production function, in its stochastic form, may be expressed as

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \tag{7.9.1}$$

where Y = output
X_2 = labor input
X_3 = capital input
u = stochastic disturbance term
e = base of natural logarithm

From Eq. (7.9.1) it is clear that the relationship between output and the two inputs is nonlinear. However, if we log-transform this model, we obtain:

[15] Arther S. Goldberger, op. cit., pp. 177–178.

$$\begin{aligned} \ln Y_i &= \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \\ &= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \end{aligned} \tag{7.9.2}$$

where $\beta_0 = \ln \beta_1$.

Thus written, the model is linear in the parameters β_0, β_2, and β_3 and is therefore a linear regression model. Notice, though, it is nonlinear in the variables Y and X but linear in the logs of these variables. In short, (7.9.2) is a *log-log, double-log,* or *log-linear model,* the multiple regression counterpart of the two-variable log-linear model (6.5.3).

The properties of the Cobb–Douglas production function are quite well known:

1. β_2 is the (partial) elasticity of output with respect to the labor input, that is, it measures the percentage change in output for, say, a 1 percent change in the labor input, holding the capital input constant (see exercise 7.9).

2. Likewise, β_3 is the (partial) elasticity of output with respect to the capital input, holding the labor input constant.

3. The sum $(\beta_2 + \beta_3)$ gives information about the *returns to scale,* that is, the response of output to a proportionate change in the inputs. If this sum is 1, then there are *constant returns to scale,* that is, doubling the inputs will double the output, tripling the inputs will triple the output, and so on. If the sum is less than 1, there are *decreasing returns to scale*—doubling the inputs will less than double the output. Finally, if the sum is greater than 1, there are *increasing returns to scale*—doubling the inputs will more than double the output.

Before proceeding further, note that whenever you have a log–linear regression model involving any number of variables the coefficient of each of the X variables measures the (partial) elasticity of the dependent variable Y with respect to that variable. Thus, if you have a k-variable log-linear model:

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \cdots + \beta_k \ln X_{ki} + u_i \tag{7.9.3}$$

each of the (partial) regression coefficients, β_2 through β_k, is the (partial) elasticity of Y with respect to variables X_2 through X_k.[16]

To illustrate the Cobb–Douglas production function, we obtained the data shown in Table 7.3; these data are for the agricultural sector of Taiwan for 1958–1972.

Assuming that the model (7.9.2) satisfies the assumptions of the classical linear regression model,[17] we obtained the following regression by the OLS

[16]To see this, differentiate (7.9.3) partially with respect to the log of each X variable. Therefore, $\partial \ln Y / \partial \ln X_2 = (\partial Y / \partial X_2)(X_2 / Y) = \beta_2$, which, by definition, is the elasticity of Y with respect to X_2 and $\partial \ln Y / \partial \ln X_3 = (\partial Y / \partial X_3)(X_3 / Y) = \beta_3$, which is the elasticity of Y with respect to X_3, and so on.

[17]Notice that in the Cobb–Douglas production function (7.9.1) we have introduced the stochastic error term in a special way so that in the resulting logarithmic transformation it enters in the usual linear form. On this, see Sec. 6.9.

TABLE 7.3 REAL GROSS PRODUCT, LABOR DAYS, AND REAL CAPITAL INPUT IN THE AGRICULTURAL SECTOR OF TAIWAN, 1958–1972

Year	Real gross product (millions of NT $)*, Y	Labor days (millions of days), X_2	Real capital input (millions of NT $), X_3
1958	16,607.7	275.5	17,803.7
1959	17,511.3	274.4	18,096.8
1960	20,171.2	269.7	18,271.8
1961	20,932.9	267.0	19,167.3
1962	20,406.0	267.8	19,647.6
1963	20,831.6	275.0	20,803.5
1964	24,806.3	283.0	22,076.6
1965	26,465.8	300.7	23,445.2
1966	27,403.0	307.5	24,939.0
1967	28,628.7	303.7	26,713.7
1968	29,904.5	304.7	29,957.8
1969	27,508.2	298.6	31,585.9
1970	29,035.5	295.5	33,474.5
1971	29,281.5	299.0	34,821.8
1972	31,535.8	288.1	41,794.3

Source: Thomas Pei-Fan Chen, "Economic Growth and Structural Change in Taiwan—1952–1972, A Production Function Approach," unpublished Ph.D. thesis, Dept. of Economics, Graduate Center, City University of New York, June 1976, Table II.

*New Taiwan dollars.

method (see Appendix 7A, Section 7A.5 for the computer printout):

$$\begin{aligned} \widehat{\ln Y_i} = \ & -3.3384 + 1.4988 \ln X_{2i} + 0.4899 \ln X_{3i} \\ & (2.4495) \quad (0.5398) \qquad (0.1020) \\ t = & (-1.3629) \quad (2.7765) \qquad (4.8005) \\ & R^2 = 0.8890 \qquad \text{df} = 12 \\ & \bar{R}^2 = 0.8705 \end{aligned} \tag{7.9.4}$$

From Eq. (7.9.4) we see that in the Taiwanese agricultural sector for the period 1958–1972 the output elasticities of labor and capital were 1.4988 and 0.4899, respectively. In other words, over the period of study, holding the capital input constant, a 1 percent increase in the labor input led on the average to about a 1.5 percent increase in the output. Similarly, holding the labor input constant, a 1 percent increase in the capital input led on the average to about a 0.5 percent increase in the output. Adding the two output elasticities, we obtain 1.9887, which gives the value of the returns to scale parameter. As is evident, over the period of the study, the Taiwanese agricultural sector was characterized by increasing returns to scale.[18]

[18]We abstain from the question of the appropriateness of the model from the theoretical viewpoint as well as the question of whether one can measure returns to scale from time series data.

From a purely statistical viewpoint, the estimated regression line fits the data quite well. The R^2 value of 0.8890 means that about 89 percent of the variation in the (log of) output is explained by the (logs of) labor and capital. In Chapter 8, we shall see how the estimated standard errors can be used to test hypotheses about the "true" values of the parameters of the Cobb–Douglas production function for the Taiwanese economy.

7.10 POLYNOMIAL REGRESSION MODELS

We now consider a class of multiple regression models, the **polynomial regression models,** that have found extensive use in econometric research relating to cost and production functions. In introducing these models, we further extend the range of models to which the classical linear regression model can easily be applied.

To fix the ideas, consider Figure 7.1, which relates the short-run marginal cost (MC) of production (Y) of a commodity to the level of its output (X). The visually-drawn MC curve in the figure, the textbook U-shaped curve, shows that the relationship between MC and output is nonlinear. If we were to quantify this relationship from the given scatterpoints, how would we go about it? In other words, what type of econometric model would capture first the declining and then the increasing nature of marginal cost?

Geometrically, the MC curve depicted in Figure 7.1 represents a *parabola*. Mathematically, the parabola is represented by the following equation:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \tag{7.10.1}$$

which is called a *quadratic function*, or more generally, a *second-degree polynomial* in the variable X—the highest power of X represents the degree of the polynomial (if X^3 were added to the preceding function, it would be a third-degree polynomial, and so on).

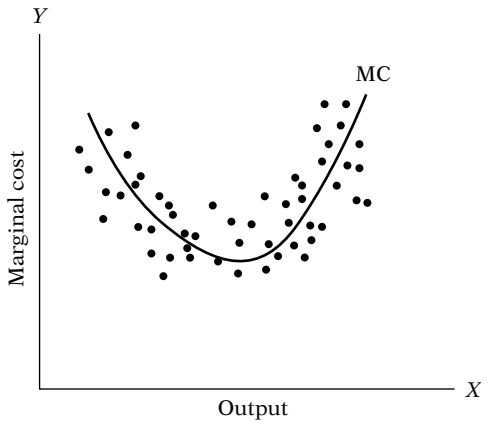

FIGURE 7.1 The U-shaped marginal cost curve.

The stochastic version of (7.10.1) may be written as

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i \qquad \textbf{(7.10.2)}$$

which is called a *second-degree polynomial* regression.

The general *kth degree polynomial regression* may be written as

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i \qquad \textbf{(7.10.3)}$$

Notice that in these types of polynomial regressions there is only one explanatory variable on the right-hand side but it appears with various powers, thus making them multiple regression models. Incidentally, note that if X_i is assumed to be fixed or nonstochastic, the powered terms of X_i also become fixed or nonstochastic.

Do these models present any special estimation problems? Since the second-degree polynomial (7.10.2) or the kth degree polynomial (7.10.13) is linear in the parameters, the β's, they can be estimated by the usual OLS or ML methodology. But what about the collinearity problem? Aren't the various X's highly correlated since they are all powers of X? Yes, but remember that terms like X^2, X^3, X^4, etc., are all nonlinear functions of X and hence, strictly speaking, do not violate the no multicollinearity assumption. In short, polynomial regression models can be estimated by the techniques presented in this chapter and present no new estimation problems.

EXAMPLE 7.4

ESTIMATING THE TOTAL COST FUNCTION

As an example of the polynomial regression, consider the data on output and total cost of production of a commodity in the short run given in Table 7.4. What type of regression model will fit these data? For this purpose, let us first draw the scattergram, which is shown in Figure 7.2.

From this figure it is clear that the relationship between total cost and output resembles the elongated S curve; notice how the total cost curve first increases gradually and then rapidly, as predicted by the celebrated law of *diminishing returns*. This S shape of the total cost curve can be captured by the following cubic or *third-degree polynomial*:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \qquad \textbf{(7.10.4)}$$

where Y = total cost and X = output.

TABLE 7.4 TOTAL COST (Y) AND OUTPUT (X)

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

Given the data of Table 7.4, we can apply the OLS method to estimate the parameters of (7.10.4). But before we do that, let us find out what economic theory has to say about the short-run cubic cost function (7.10.4). Elementary price theory shows that in the short run the

(*Continued*)

EXAMPLE 7.4 (*Continued*)

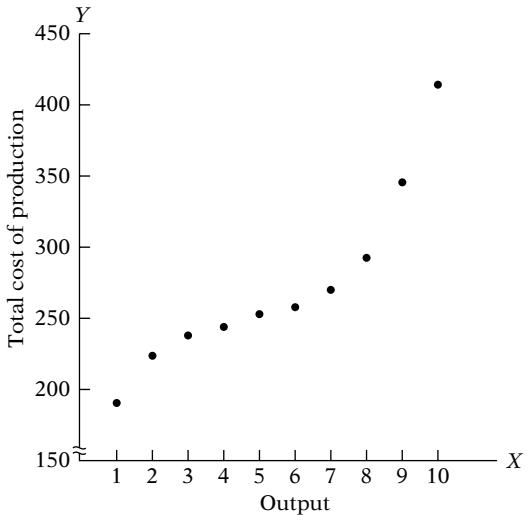

FIGURE 7.2 The total cost curve.

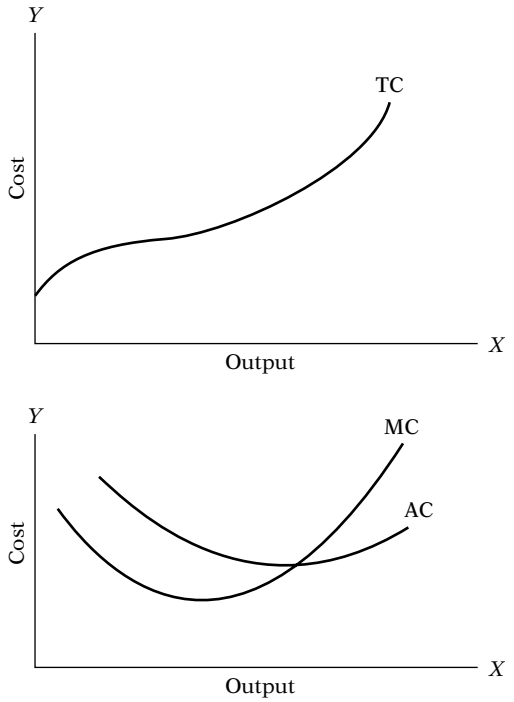

FIGURE 7.3 Short-run cost functions.

marginal cost (MC) and average cost (AC) curves of production are typically U-shaped—initially, as output increases both MC and AC decline, but after a certain level of output they both turn upward, again the consequence of the law of diminishing return. This can be seen in Figure 7.3 (see also Figure 7.1). And since the MC and AC curves are derived from the total cost curve, the U-shaped nature of these curves puts some restrictions on the parameters of the total cost curve (7.10.4). As a matter of fact, it can be shown that the parameters of (7.10.4) must satisfy the following restrictions if one is to observe the typical U-shaped short-run marginal and average cost curves:[19]

$$\begin{aligned} &\textbf{1. } \beta_0, \beta_1, \text{ and } \beta_3 > 0 \\ &\textbf{2. } \beta_2 < 0 \\ &\textbf{3. } \beta_2^2 < 3\beta_1\beta_3 \end{aligned} \qquad \textbf{(7.10.5)}$$

All this theoretical discussion might seem a bit tedious. But this knowledge is extremely useful when we examine the empirical results, for if the empirical results do not agree with prior expectations, then, assuming we have not committed a specification error (i.e., chosen the wrong model), we will have to modify our theory or look for a new theory and start the empirical enquiry all over again. But as noted in the Introduction, this is the nature of any empirical investigation.

Empirical Results

When the third-degree polynomial regression was fitted to the data of Table 7.4, we obtained the following results:

$$\begin{aligned} \hat{Y}_i = 141.7667 &+ 63.4776X_i - 12.9615X_i^2 + 0.9396X_i^3 \\ (6.3753) \quad &(4.7786) \quad (0.9857) \quad (0.0591) \\ R^2 &= 0.9983 \end{aligned} \qquad \textbf{(7.10.6)}$$

(*Continued*)

[19]See Alpha C. Chiang, *Fundamental Methods of Mathematical Economics,* 3d ed., McGraw-Hill, New York, 1984, pp. 250–252.

EXAMPLE 7.4 (*Continued*)

(*Note:* The figures in parentheses are the estimated standard errors.) Although we will examine the statistical significance of these results in the next chapter, the reader can verify that they are in conformity with the theoretical expectations listed in (7.10.5). We leave it as an exercise for the reader to interpret the regression (7.10.6).

EXAMPLE 7.5

GDP GROWTH RATE, 1960–1985 AND RELATIVE PER CAPITA GDP, IN 119 DEVELOPING COUNTRIES

As an additional economic example of the polynomial regression model, consider the following regression results[20]:

$$\widehat{\text{GDPG}}_i = 0.013 + 0.062\,\text{RGDP} - 0.061\,\text{RGDP}^2 \quad \textbf{(7.10.7)}$$

$$\text{se} = (0.004) \quad (0.027) \quad (0.033)$$

$$R^2 = 0.053 \qquad \text{adj } R^2 = 0.036$$

where GDPG = GDP growth rate, percent (average for 1960–1985), and RGDP = relative per capita GDP, 1960 (percentage of U.S. GDP per capita, 1960). The adjusted R^2 (adj R^2) tells us that, after taking into account the number of regressors, the model explains only about 3.6 percent of the variation in GDPG. Even the unadjusted R^2 of 0.053 seems low. This might sound a disappointing value but, as we shall show in the next chapter, such low R^2's are frequently encountered in cross-sectional data with a large number of observations. Besides, even an apparently low R^2 value can be statistically significant (i.e., different from zero), as we will show in the next chapter.

As this regression shows, GDPG in developing countries increased as RGDP increased, but at a decreasing rate; that is, developing economies were not catching up with advanced economies.[21] This example shows how relatively simple econometric models can be used to shed light on important economic phenomena.

*7.11 PARTIAL CORRELATION COEFFICIENTS

Explanation of Simple and Partial Correlation Coefficients

In Chapter 3 we introduced the coefficient of correlation r as a measure of the degree of linear association between two variables. For the three-variable

[20]Source: *The East Asian Economic Miracle: Economic Growth and Public Policy*, A World Bank Policy Research Report, Oxford University Press, U.K, 1993, p. 29.

[21]If you take the derivative of (7.10.7), you will obtain

$$\frac{\text{dGDPG}}{\text{dRGDP}} = 0.062 - 0.122\,\text{RGDP}$$

showing that the rate of change of GDPG with respect to RGDP is declining. If you set this derivative to zero, you will get RGDP $\approx$ 0.5082. Thus, if a country's GDP reaches about 51 percent of the U.S. GDP, the *rate of growth* of GDPG will crawl to zero.

*Optional.

regression model we can compute three correlation coefficients: r_{12} (correlation between Y and X_2), r_{13} (correlation coefficient between Y and X_3), and r_{23} (correlation coefficient between X_2 and X_3); notice that we are letting the subscript 1 represent Y for notational convenience. These correlation coefficients are called **gross** or **simple correlation coefficients,** or **correlation coefficients of zero order.** These coefficients can be computed by the definition of correlation coefficient given in (3.5.13).

But now consider this question: Does, say, r_{12} in fact measure the "true" degree of (linear) association between Y and X_2 when a third variable X_3 may be associated with both of them? This question is analogous to the following question: Suppose the true regression model is (7.1.1) but we omit from the model the variable X_3 and simply regress Y on X_2, obtaining the slope coefficient of, say, b_{12}. Will this coefficient be equal to the true coefficient β_2 if the model (7.1.1) were estimated to begin with? The answer should be apparent from our discussion in Section 7.7. In general, r_{12} is not likely to reflect the true degree of association between Y and X_2 in the presence of X_3. As a matter of fact, it is likely to give a false impression of the nature of association between Y and X_2, as will be shown shortly. Therefore, what we need is a correlation coefficient that is independent of the influence, if any, of X_3 on X_2 and Y. Such a correlation coefficient can be obtained and is known appropriately as the **partial correlation coefficient.** Conceptually, it is similar to the partial regression coefficient. We define

$r_{12.3}$ = partial correlation coefficient between Y and X_2, holding X_3 constant

$r_{13.2}$ = partial correlation coefficient between Y and X_3, holding X_2 constant

$r_{23.1}$ = partial correlation coefficient between X_2 and X_3, holding Y constant

These partial correlations can be easily obtained from the simple or zero-order, correlation coefficients as follows (for proofs, see the exercises)[22]:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \tag{7.11.1}$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}} \tag{7.11.2}$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}} \tag{7.11.3}$$

The partial correlations given in Eqs. (7.11.1) to (7.11.3) are called **first-order correlation coefficients.** By *order* we mean the number of secondary

[22]Most computer programs for multiple regression analysis routinely compute the simple correlation coefficients; hence the partial correlation coefficients can be readily computed.

subscripts. Thus $r_{12.34}$ would be the correlation coefficient of order two, $r_{12.345}$ would be the correlation coefficient of order three, and so on. As noted previously, r_{12}, r_{13}, and so on are called *simple* or *zero-order correlations*. The interpretation of, say, $r_{12.34}$ is that it gives the coefficient of correlation between Y and X_2, holding X_3 and X_4 constant.

Interpretation of Simple and Partial Correlation Coefficients

In the two-variable case, the simple r had a straightforward meaning: It measured the degree of (linear) association (and not causation) between the dependent variable Y and the single explanatory variable X. But once we go beyond the two-variable case, we need to pay careful attention to the interpretation of the simple correlation coefficient. From (7.11.1), for example, we observe the following:

1. Even if $r_{12} = 0$, $r_{12.3}$ will not be zero unless r_{13} or r_{23} or both are zero.

2. If $r_{12} = 0$ and r_{13} and r_{23} are nonzero and are of the same sign, $r_{12.3}$ will be negative, whereas if they are of the opposite signs, it will be positive. An example will make this point clear. Let Y = crop yield, X_2 = rainfall, and X_3 = temperature. Assume $r_{12} = 0$, that is, no association between crop yield and rainfall. Assume further that r_{13} is positive and r_{23} is negative. Then, as (7.11.1) shows, $r_{12.3}$ will be positive; that is, holding temperature constant, there is a positive association between yield and rainfall. This seemingly paradoxical result, however, is not surprising. Since temperature X_3 affects both yield Y and rainfall X_2, in order to find out the net relationship between crop yield and rainfall, we need to remove the influence of the "nuisance" variable temperature. This example shows how one might be misled by the simple coefficient of correlation.

3. The terms $r_{12.3}$ and r_{12} (and similar comparisons) need not have the same sign.

4. In the two-variable case we have seen that r^2 lies between 0 and 1. The same property holds true of the squared partial correlation coefficients. Using this fact, the reader should verify that one can obtain the following expression from (7.11.1):

$$0 \leq r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1 \qquad \textbf{(7.11.4)}$$

which gives the interrelationships among the three zero-order correlation coefficients. Similar expressions can be derived from Eqs. (7.9.3) and (7.9.4).

5. Suppose that $r_{13} = r_{23} = 0$. Does this mean that r_{12} is also zero? The answer is obvious from (7.11.4). The fact that Y and X_3 and X_2 and X_3 are uncorrelated does not mean that Y and X_2 are uncorrelated.

In passing, note that the expression $r_{12.3}^2$ may be called the **coefficient of partial determination** and may be interpreted as the proportion of the variation in Y not explained by the variable X_3 that has been explained

by the inclusion of X_2 into the model (see exercise 7.5). Conceptually it is similar to R^2.

Before moving on, note the following relationships between R^2, simple correlation coefficients, and partial correlation coefficients:

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \tag{7.11.5}$$

$$R^2 = r_{12}^2 + (1 - r_{12}^2)\, r_{13.2}^2 \tag{7.11.6}$$

$$R^2 = r_{13}^2 + (1 - r_{13}^2)\, r_{12.3}^2 \tag{7.11.7}$$

In concluding this section, consider the following: It was stated previously that R^2 will not decrease if an additional explanatory variable is introduced into the model, which can be seen clearly from (7.11.6). This equation states that the proportion of the variation in Y explained by X_2 and X_3 jointly is the sum of two parts: the part explained by X_2 alone ($= r_{12}^2$) and the part not explained by X_2 ($= 1 - r_{12}^2$) times the proportion that is explained by X_3 after holding the influence of X_2 constant. Now $R^2 > r_{12}^2$ so long as $r_{13.2}^2 > 0$. At worst, $r_{13.2}^2$ will be zero, in which case $R^2 = r_{12}^2$.

7.12 SUMMARY AND CONCLUSIONS

1. This chapter introduced the simplest possible multiple linear regression model, namely, the three-variable regression model. It is understood that the term *linear* refers to linearity in the parameters and not necessarily in the variables.

2. Although a three-variable regression model is in many ways an extension of the two-variable model, there are some new concepts involved, such as *partial regression coefficients, partial correlation coefficients, multiple correlation coefficient, adjusted and unadjusted (for degrees of freedom) R^2, multicollinearity,* and *specification bias*.

3. This chapter also considered the functional form of the multiple regression model, such as the *Cobb–Douglas production function* and the *polynomial regression model*.

4. Although R^2 and adjusted R^2 are overall measures of how the chosen model fits a given set of data, their importance should not be overplayed. What is critical is the underlying theoretical expectations about the model in terms of a priori signs of the coefficients of the variables entering the model and, as it is shown in the following chapter, their statistical significance.

5. The results presented in this chapter can be easily generalized to a multiple linear regression model involving any number of regressors. But the algebra becomes very tedious. This tedium can be avoided by resorting to matrix algebra. For the interested reader, the extension to the k-variable