6

EXTENSIONS OF THE TWO-VARIABLE LINEAR REGRESSION MODEL

Some aspects of linear regression analysis can be easily introduced within the framework of the two-variable linear regression model that we have been discussing so far. First we consider the case of **regression through the origin,** that is, a situation where the intercept term, β_1, is absent from the model. Then we consider the question of the **units of measurement,** that is, how the Y and X variables are measured and whether a change in the units of measurement affects the regression results. Finally, we consider the question of the **functional form** of the linear regression model. So far we have considered models that are linear in the parameters as well as in the variables. But recall that the regression theory developed in the previous chapters requires only that the parameters be linear; the variables may or may not enter linearly in the model. By considering models that are linear in the parameters but not necessarily in the variables, we show in this chapter how the two-variable models can deal with some interesting practical problems.

Once the ideas introduced in this chapter are grasped, their extension to multiple regression models is quite straightforward, as we shall show in Chapters 7 and 8.

6.1 REGRESSION THROUGH THE ORIGIN

There are occasions when the two-variable PRF assumes the following form:

$$Y_i = \beta_2 X_i + u_i \tag{6.1.1}$$

In this model the intercept term is absent or zero, hence the name **regression through the origin.**

As an illustration, consider the Capital Asset Pricing Model (CAPM) of modern portfolio theory, which, in its risk-premium form, may be expressed as[1]

$$(\text{ER}_i - r_f) = \beta_i(\text{ER}_m - r_f) \tag{6.1.2}$$

where ER_i = expected rate of return on security i
ER_m = expected rate of return on the market portfolio as represented by, say, the S&P 500 composite stock index
r_f = risk-free rate of return, say, the return on 90-day Treasury bills
β_i = the Beta coefficient, a measure of systematic risk, i.e., risk that cannot be eliminated through diversification. Also, a measure of the extent to which the ith security's rate of return moves with the market. A $\beta_i > 1$ implies a volatile or aggressive security, whereas a $\beta_i < 1$ a defensive security. (*Note:* Do not confuse this β_i with the slope coefficient of the two-variable regression, β_2.)

If capital markets work efficiently, then CAPM postulates that security i's expected risk premium ($= \text{ER}_i - r_f$) is equal to that security's β coefficient times the expected market risk premium ($= \text{ER}_m - r_f$). If the CAPM holds, we have the situation depicted in Figure 6.1. The line shown in the figure is known as the **security market line** (SML).

For empirical purposes, (6.1.2) is often expressed as

$$R_i - r_f = \beta_i(R_m - r_f) + u_i \tag{6.1.3}$$

or

$$R_i - r_f = \alpha_i + \beta_i(R_m - r_f) + u_i \tag{6.1.4}$$

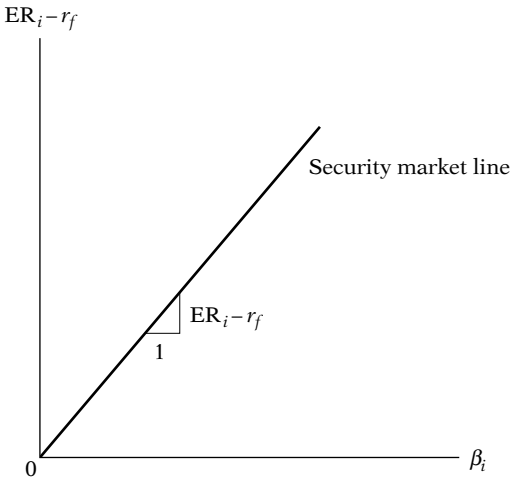

FIGURE 6.1 Systematic risk.

[1]See Haim Levy and Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice-Hall International, Englewood Cliffs, N.J., 1984, Chap. 14.

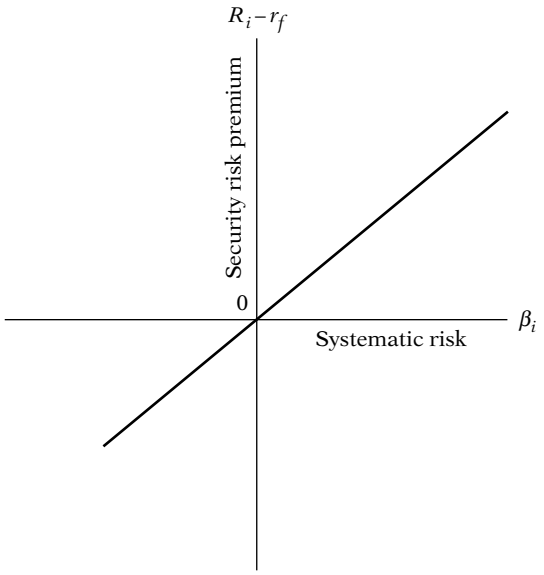

FIGURE 6.2 The Market Model of Portfolio Theory (assuming $\alpha_i = 0$).

The latter model is known as the **Market Model.**[2] If CAPM holds, α_i is expected to be zero. (See Figure 6.2.)

In passing, note that in (6.1.4) the dependent variable, Y, is $(R_i - r_f)$ and the explanatory variable, X, is β_i, the volatility coefficient, and *not* $(R_m - r_f)$. Therefore, to run regression (6.1.4), one must first estimate β_i, which is usually derived from the **characteristic line,** as described in exercise 5.5. (For further details, see exercise 8.28.)

As this example shows, sometimes the underlying theory dictates that the intercept term be absent from the model. Other instances where the zero-intercept model may be appropriate are Milton Friedman's permanent income hypothesis, which states that permanent consumption is proportional to permanent income; cost analysis theory, where it is postulated that the variable cost of production is proportional to output; and some versions of monetarist theory that state that the rate of change of prices (i.e., the rate of inflation) is proportional to the rate of change of the money supply.

How do we estimate models like (6.1.1), and what special problems do they pose? To answer these questions, let us first write the SRF of (6.1.1), namely,

$$Y_i = \hat{\beta}_2 X_i + \hat{u}_i \qquad \textbf{(6.1.5)}$$

Now applying the OLS method to (6.1.5), we obtain the following formulas for $\hat{\beta}_2$ and its variance (proofs are given in Appendix 6A, Section 6A.1):

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad \textbf{(6.1.6)}$$

[2]See, for instance, Diana R. Harrington, *Modern Portfolio Theory and the Capital Asset Pricing Model: A User's Guide*, Prentice Hall, Englewood Cliffs, N.J., 1983, p. 71.

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2} \tag{6.1.7}$$

where σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1} \tag{6.1.8}$$

It is interesting to compare these formulas with those obtained when the intercept term is included in the model:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{3.1.6}$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{3.3.1}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{3.3.5}$$

The differences between the two sets of formulas should be obvious: In the model with the intercept term absent, we use **raw** sums of squares and cross products but in the intercept-present model, we use adjusted (from mean) sums of squares and cross products. Second, the df for computing $\hat{\sigma}^2$ is $(n-1)$ in the first case and $(n-2)$ in the second case. (Why?)

Although the interceptless or zero intercept model may be appropriate on occasions, there are some features of this model that need to be noted. First, $\sum \hat{u}_i$, which is always zero for the model with the intercept term (the conventional model), need not be zero when that term is absent. In short, $\sum \hat{u}_i$ need not be zero for the regression through the origin. Second, r^2, the coefficient of determination introduced in Chapter 3, which is always nonnegative for the conventional model, can on occasions turn out to be *negative* for the interceptless model! This anomalous result arises because the r^2 introduced in Chapter 3 explicitly assumes that the intercept is included in the model. Therefore, the conventionally computed r^2 may not be appropriate for regression-through-the-origin models.[3]

r^2 for Regression-through-Origin Model

As just noted, and as further discussed in Appendix 6A, Section 6A.1, the conventional r^2 given in Chapter 3 is not appropriate for regressions that do not contain the intercept. But one can compute what is known as the **raw r^2** for such models, which is defined as

$$\text{raw } r^2 = \frac{\left(\sum X_i Y_i\right)^2}{\sum X_i^2 \sum Y_i^2} \tag{6.1.9}$$

[3]For additional discussion, see Dennis J. Aigner, *Basic Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1971, pp. 85–88.

Note: These are raw (i.e., not mean-corrected) sums of squares and cross products.

Although this raw r^2 satisfies the relation $0 < r^2 < 1$, it is not directly comparable to the conventional r^2 value. For this reason some authors do not report the r^2 value for zero intercept regression models.

Because of these special features of this model, one needs to exercise great caution in using the zero intercept regression model. *Unless there is very strong a priori expectation*, one would be well advised to stick to the conventional, intercept-present model. This has a dual advantage. First, if the intercept term is included in the model but it turns out to be statistically insignificant (i.e., statistically equal to zero), for all practical purposes we have a regression through the origin.[4] Second, and more important, if in fact there is an intercept in the model but we insist on fitting a regression through the origin, we would be committing a **specification error,** thus violating Assumption 9 of the classical linear regression model.

AN ILLUSTRATIVE EXAMPLE:
THE CHARACTERISTIC LINE OF PORTFOLIO THEORY

Table 6.1 gives data on the annual rates of return (%) on Afuture Fund, a mutual fund whose primary investment objective is maximum capital gain, and on the market portfolio, as measured by the Fisher Index, for the period 1971–1980.

In exercise 5.5 we introduced the *characteristic line* of investment analysis, which can be written as

$$Y_i = \alpha_i + \beta_i X_i + u_i \qquad \textbf{(6.1.10)}$$

where Y_i = annual rate of return (%) on Afuture Fund
X_i = annual rate of return (%) on the market portfolio
β_i = slope coefficient, also known as the **Beta** coefficient in portfolio theory, and
α_i = the intercept

In the literature there is no consensus about the prior value of α_i. Some empirical results have shown it to be positive and statistically significant and some have shown it to be not statistically significantly different from zero; in the latter case we could write the model as

$$Y_i = \beta_i X_i + u_i \qquad \textbf{(6.1.11)}$$

that is, a regression through the origin.

TABLE 6.1
ANNUAL RATES OF RETURN ON AFUTURE FUND AND ON THE FISHER INDEX (MARKET PORTFOLIO), 1971–1980

Year	Return on Afuture Fund, % *Y*	Return on Fisher Index, % *X*
1971	67.5	19.5
1972	19.2	8.5
1973	−35.2	−29.3
1974	−42.0	−26.5
1975	63.7	61.9
1976	19.3	45.5
1977	3.6	9.5
1978	20.0	14.0
1979	40.3	35.3
1980	37.5	31.0

Source: Haim Levy and Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice,* Prentice-Hall International, Englewood Cliffs, N.J., 1984, pp. 730 and 738. These data were obtained by the authors from Weisenberg Investment Service, *Investment Companies,* 1981 edition.

(*Continued*)

[4]Henri Theil points out that if the intercept is in fact absent, the slope coefficient may be estimated with far greater precision than with the intercept term left in. See his *Introduction to Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1978, p. 76. See also the numerical example given next.

AN ILLUSTRATIVE EXAMPLE (*Continued*)

If we decide to use model (6.1.11), we obtain the following regression results

$$\hat{Y}_i = 1.0899\, X_i \qquad \textbf{(6.1.12)}$$
$$(0.1916) \qquad \text{raw } r^2 = 0.7825$$
$$t = (5.6884)$$

which shows that β_i is significantly greater than zero. The interpretation is that a 1 percent increase in the market rate of return leads on the average to about 1.09 percent increase in the rate of return on Afuture Fund.

How can we be sure that model (6.1.11), not (6.1.10), is appropriate, especially in view of the fact that there is no strong a priori belief in the hypothesis that α_i is in fact zero? This can be checked by running the regression (6.1.10). Using the data given in Table 6.1, we obtained the following results:

$$\hat{Y}_i = 1.2797 + 1.0691 X_i \qquad \textbf{(6.1.13)}$$
$$(7.6886) \quad (0.2383)$$
$$t = (0.1664) \quad (4.4860) \qquad r^2 = 0.7155$$

Note: The r^2 values of (6.1.12) and (6.1.13) are *not* directly comparable. From these results one cannot reject the hypothesis that the true intercept is equal to zero, thereby justifying the use of (6.1.1), that is, regression through the origin.

In passing, note that there is not a great deal of difference in the results of (6.1.12) and (6.1.13), although the estimated standard error of $\hat{\beta}$ is slightly lower for the regression-through-the-origin model, thus supporting Theil's argument given in footnote 4 that if α_i is in fact zero, the slope coefficient may be measured with greater precision: using the data given in Table 6.1 and the regression results, the reader can easily verify that the 95% confidence interval for the slope coefficient of the regression-through-the-origin model is (0.6566, 1.5232) whereas for the model (6.1.13) it is (0.5195, 1.6186); that is, the former confidence interval is narrower that the latter.

6.2 SCALING AND UNITS OF MEASUREMENT

To grasp the ideas developed in this section, consider the data given in Table 6.2, which refers to U.S. gross private domestic investment (GPDI) and gross domestic product (GDP), in billions as well as millions of (chained) 1992 dollars.

TABLE 6.2 GROSS PRIVATE DOMESTIC INVESTMENT AND GDP, UNITED STATES, 1988–1997

Observation	GPDIBL	GPDIM	GDPB	GDPM
1988	828.2000	828200.0	5865.200	5865200
1989	863.5000	863500.0	6062.000	6062000
1990	815.0000	815000.0	6136.300	6136300
1991	738.1000	738100.0	6079.400	6079400
1992	790.4000	790400.0	6244.400	6244400
1993	863.6000	863600.0	6389.600	6389600
1994	975.7000	975700.0	6610.700	6610700
1995	996.1000	996100.0	6761.600	6761600
1996	1084.1000	1084100.0	6994.800	6994800
1997	1206.4000	1206400.0	7269.800	7269800

Note: GPDIBL = gross private domestic investment, billions of 1992 dollars.
GPDIM = gross private domestic investments, millions of 1992 dollars.
GDPB = gross domestic product, billions of 1992 dollars.
GDPM = gross domestic product, millions of 1992 dollars.
Source: Economic Report of the President, 1999, Table B-2, p. 328.

Suppose in the regression of GPDI on GDP one researcher uses data in billions of dollars but another expresses data in millions of dollars. Will the regression results be the same in both cases? If not, which results should one use? In short, do the units in which the regressand and regressor(s) are measured make any difference in the regression results? If so, what is the sensible course to follow in choosing units of measurement for regression analysis? To answer these questions, let us proceed systematically. Let

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{6.2.1}$$

where Y = GPDI and X = GDP. Define

$$Y_i^* = w_1 Y_i \tag{6.2.2}$$

$$X_i^* = w_2 X_i \tag{6.2.3}$$

where w_1 and w_2 are constants, called the **scale factors;** w_1 may equal w_2 or be different.

From (6.2.2) and (6.2.3) it is clear that Y_i^* and X_i^* are *rescaled* Y_i and X_i. Thus, if Y_i and X_i are measured in billions of dollars and one wants to express them in millions of dollars, we will have $Y_i^* = 1000\, Y_i$ and $X_i^* = 1000\, X_i$; here $w_1 = w_2 = 1000$.

Now consider the regression using Y_i^* and X_i^* variables:

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \tag{6.2.4}$$

where $Y_i^* = w_1 Y_i$, $X_i^* = w_2 X_i$, and $\hat{u}_i^* = w_1 \hat{u}_i$. (Why?)

We want to find out the relationships between the following pairs:

1. $\hat{\beta}_1$ and $\hat{\beta}_1^*$
2. $\hat{\beta}_2$ and $\hat{\beta}_2^*$
3. $\text{var}\,(\hat{\beta}_1)$ and $\text{var}\,(\hat{\beta}_1^*)$
4. $\text{var}\,(\hat{\beta}_2)$ and $\text{var}\,(\hat{\beta}_2^*)$
5. $\hat{\sigma}^2$ and $\hat{\sigma}^{*2}$
6. r_{xy}^2 and $r_{x^*y^*}^2$

From least-squares theory we know (see Chapter 3) that

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \tag{6.2.5}$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{6.2.6}$$

$$\text{var}\,(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \sigma^2 \tag{6.2.7}$$

$$\text{var}\,(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{6.2.8}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{6.2.9}$$

Applying the OLS method to (6.2.4), we obtain similarly

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \tag{6.2.10}$$

$$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} \tag{6.2.11}$$

$$\text{var}\,(\hat{\beta}_1^*) = \frac{\sum X_i^{*2}}{n \sum x_i^{*2}} \cdot \sigma^{*2} \tag{6.2.12}$$

$$\text{var}\,(\hat{\beta}_2^*) = \frac{\sigma^{*2}}{\sum x_i^{*2}} \tag{6.2.13}$$

$$\hat{\sigma}^{*2} = \frac{\sum \hat{u}_i^{*2}}{(n-2)} \tag{6.2.14}$$

From these results it is easy to establish relationships between the two sets of parameter estimates. All that one has to do is recall these definitional relationships: $Y_i^* = w_1 Y_i$ (or $y_i^* = w_1 y_i$); $X_i^* = w_2 X_i$ (or $x_i^* = w_2 x_i$); $\hat{u}_i^* = w_1 \hat{u}_i$; $\bar{Y}^* = w_1 \bar{Y}$ and $\bar{X}^* = w_2 \bar{X}$. Making use of these definitions, the reader can easily verify that

$$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right) \hat{\beta}_2 \tag{6.2.15}$$

$$\hat{\beta}_1^* = w_1 \hat{\beta}_1 \tag{6.2.16}$$

$$\hat{\sigma}^{*2} = w_1^2 \hat{\sigma}^2 \tag{6.2.17}$$

$$\text{var}\,(\hat{\beta}_1^*) = w_1^2\, \text{var}\,(\hat{\beta}_1) \tag{6.2.18}$$

$$\text{var}\,(\hat{\beta}_2^*) = \left(\frac{w_1}{w_2}\right)^2 \text{var}\,(\hat{\beta}_2) \tag{6.2.19}$$

$$r_{xy}^2 = r_{x^* y^*}^2 \tag{6.2.20}$$

From the preceding results it should be clear that, given the regression results based on one scale of measurement, one can derive the results based on another scale of measurement once the scaling factors, the *w*'s, are known. In practice, though, one should choose the units of measurement sensibly; there is little point in carrying all those zeros in expressing numbers in millions or billions of dollars.

From the results given in (6.2.15) through (6.2.20) one can easily derive some special cases. For instance, if $w_1 = w_2$, that is, the scaling factors are identical, the slope coefficient and its standard error remain unaffected in going from the (Y_i, X_i) to the (Y_i^*, X_i^*) scale, which should be intuitively clear. However, the intercept and its standard error are both multiplied by w_1. But if the X scale is not changed (i.e., $w_2 = 1$) and the Y scale is changed by the factor w_1, the slope as well as the intercept coefficients and their respective standard errors are all multiplied by the same w_1 factor. Finally, if the Y scale remains unchanged (i.e., $w_1 = 1$) but the X scale is changed by the factor w_2, the slope coefficient and its standard error are multiplied by the factor $(1/w_2)$ but the intercept coefficient and its standard error remain unaffected.

It should, however, be noted that the transformation from the (Y, X) to the (Y^*, X^*) scale does not affect the properties of the OLS estimators discussed in the preceding chapters.

A NUMERICAL EXAMPLE: THE RELATIONSHIP BETWEEN GPDI AND GDP, UNITED STATES, 1988–1997

To substantiate the preceding theoretical results, let us return to the data given in Table 6.2 and examine the following results (numbers in parentheses are the estimated standard errors).

Both GPDI and GDP in billions of dollars:

$$\widehat{\text{GPDI}}_t = -1026.498 + 0.3016\ \text{GDP}_t \qquad \text{(6.2.21)}$$
$$\text{se} = (257.5874)\quad (0.0399)\qquad r^2 = 0.8772$$

Both GPDI and GDP in millions of dollars:

$$\widehat{\text{GPDI}}_t = -1{,}026{,}498 + 0.3016\ \text{GDP}_t \qquad \text{(6.2.22)}$$
$$\text{se} = (257{,}587.4)\quad (0.0399)\qquad r^2 = 0.8772$$

Notice that the intercept as well as its standard error is 1000 times the corresponding values in the regression (6.2.21) (note that $w_1 = 1000$ in going from billions to millions of dollars), but the slope coefficient as well as its standard error is unchanged, in accordance with theory.

GPDI in billions of dollars and GDP in millions of dollars:

$$\widehat{\text{GPDI}}_t = -1026.498 + 0.000301\ \text{GDP}_t \qquad \text{(6.2.23)}$$
$$\text{se} = (257.5874)\quad (0.0000399)\qquad r^2 = 0.8772$$

As expected, the slope coefficient as well as its standard error is 1/1000 its value in (6.2.21), since only the X, or GDP, scale is changed.

GPDI in millions of dollars and GDP in billions of dollars:

$$\widehat{\text{GPDI}}_t = -1{,}026{,}498 + 301.5826\ \text{GDP}_t \qquad \text{(6.2.24)}$$
$$\text{se} = (257{,}587.4)\quad (39.89989)\qquad r^2 = 0.8772$$

Again notice that both the intercept and the slope coefficients as well as their respective standard errors are 1000 times their values in (6.2.21), in accordance with our theoretical results.

Notice that in all the regressions presented above the r^2 value remains the same, which is not surprising because the r^2 value is *invariant* to changes in the unit of measurement, as it is a pure, or dimensionless, number.

A Word about Interpretation

Since the slope coefficient β_2 is simply the rate of change, it is measured in the units of the ratio

$$\frac{\text{Units of the dependent variable}}{\text{Units of the explanatory variable}}$$

Thus in regression (6.2.21) the interpretation of the slope coefficient 0.3016 is that if GDP changes by a unit, which is 1 billion dollars, GPDI on the average changes by 0.3016 billion dollars. In regression (6.2.23) a unit change in GDP, which is 1 million dollars, leads on average to a 0.000302 billion dollar change in GPDI. The two results are of course identical in the effects of GDP on GPDI; they are simply expressed in different units of measurement.

6.3 REGRESSION ON STANDARDIZED VARIABLES

We saw in the previous section that the units in which the regressand and regressor(s) are expressed affect the interpretation of the regression coefficients. This can be avoided if we are willing to express the regressand and regressor(s) as *standardized variables*. A variable is said to be standardized if we subtract the mean value of the variable from its individual values and divide the difference by the standard deviation of that variable.

Thus, in the regression of Y and X, if we redefine these variables as

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \tag{6.3.1}$$

$$X_i^* = \frac{X_i - \bar{X}}{S_X} \tag{6.3.2}$$

where $\bar{Y}$ = sample mean of Y, S_Y = sample standard deviation of Y, $\bar{X}$ = sample mean of X, and S_X is the sample standard deviation of X; the variables Y_i^* and X_i^* are called **standardized variables.**

An interesting property of a standardized variable is that its mean value is always zero and its standard deviation is always 1. (For proof, see Appendix 6A, Section 6A.2.)

As a result, it does not matter in what unit the regressand and regressor(s) are measured. Therefore, instead of running the standard (bivariate) regression:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{6.3.3}$$

we could run regression on the standardized variables as

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + u_i^* \tag{6.3.4}$$

$$= \beta_2^* X_i^* + u_i^* \tag{6.3.5}$$

since it is easy to show that, in the regression involving standardized regressand and regressor(s), the intercept term is always zero.[5] The regression coefficients of the standardized variables, denoted by β_1^* and β_2^*, are known in the literature as the **beta coefficients.**[6] Incidentally, notice that (6.3.5) is a regression through the origin.

How do we interpret the beta coefficients? The interpretation is that if the (standardized) regressor increases by one standard deviation, on average, the (standardized) regressand increases by β_2^* standard deviation units. Thus, unlike the traditional model (6.3.3), we measure the effect not in terms of the original units in which Y and X are expressed, but in standard deviation units.

To show the difference between (6.3.3) and (6.3.5), let us return to the GPDI and GDP example discussed in the preceding section. The results of (6.2.21) discussed previously are reproduced here for convenience.

$$\begin{aligned} \widehat{\text{GPDI}}_t &= -1026.498 \quad + \quad 0.3016\ \text{GDP}_t \\ \text{se} &= \quad (257.5874) \quad (0.0399) \qquad r^2 = 0.8872 \end{aligned} \tag{6.3.6}$$

where GPDI and GDP are measured in billions of dollars.

The results corresponding to (6.3.5) are as follows, where the starred variables are standardized variables:

$$\begin{aligned} \widehat{\text{GPDI}}_t^* &= \ 0.9387\ \text{GDP}_t^* \\ \text{se} &= (0.1149) \end{aligned} \tag{6.3.7}$$

We know how to interpret (6.3.6): If GDP goes up by a dollar, on average GPDI goes up by about 30 cents. How about (6.3.7)? Here the interpretation is that if the (standardized) GDP increases by one standard deviation, on average, the (standardized) GPDI increases by about 0.94 standard deviations.

What is the advantage of the standardized regression model over the traditional model? The advantage becomes more apparent if there is more than one regressor, a topic we will take up in Chapter 7. By standardizing all regressors, we put them on equal basis and therefore can compare them directly. If the coefficient of a standardized regressor is larger than that of another standardized regressor appearing in that model, then the latter contributes more relatively to the explanation of the regressand than the latter. In other words, we can use the beta coefficients as a measure of relative strength of the various regressors. But more on this in the next two chapters.

Before we leave this topic, two points may be noted. First, for the standardized regression (6.3.7) we have not given the r^2 value because this is a regression through the origin for which the usual r^2 is not applicable, as pointed out in Section 6.1. Second, there is an interesting relationship between the β coefficients of the conventional model and the beta coefficients.

[5]Recall from Eq. (3.1.7) that intercept = mean value of the dependent variable − slope times the mean value of the regressor. But for the standardized variables the mean values of the dependent variable and the regressor are zero. Hence the intercept value is zero.

[6]Do not confuse these beta coefficients with the beta coefficients of finance theory.

For the bivariate case, the relationship is as follows:

$$\hat{\beta}_2^* = \hat{\beta}_2\left(\frac{S_x}{S_y}\right) \tag{6.3.8}$$

where S_x = the sample standard deviation of the X regressor and S_y = the sample standard deviation of the regressand. Therefore, one can crisscross between the β and beta coefficients if we know the (sample) standard deviation of the regressor and regressand. We will see in the next chapter that this relationship holds true in the multiple regression also. It is left as an exercise for the reader to verify (6.3.8) for our illustrative example.

6.4 FUNCTIONAL FORMS OF REGRESSION MODELS

As noted in Chapter 2, this text is concerned primarily with models that are linear in the parameters; they may or may not be linear in the variables. In the sections that follow we consider some commonly used regression models that may be nonlinear in the variables but are linear in the parameters or that can be made so by suitable transformations of the variables. In particular, we discuss the following regression models:

1. The log-linear model
2. Semilog models
3. Reciprocal models
4. The logarithmic reciprocal model

We discuss the special features of each model, when they are appropriate, and how they are estimated. Each model is illustrated with suitable examples.

6.5 HOW TO MEASURE ELASTICITY: THE LOG-LINEAR MODEL

Consider the following model, known as the **exponential regression model:**

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \tag{6.5.1}$$

which may be expressed alternatively as[7]

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i \tag{6.5.2}$$

where ln = natural log (i.e., log to the base e, and where $e = 2.718$).[8]

If we write (6.5.2) as

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \tag{6.5.3}$$

[7]Note these properties of the logarithms: (1) $\ln(AB) = \ln A + \ln B$, (2) $\ln(A/B) = \ln A - \ln B$, and (3) $\ln(A^k) = k \ln A$, assuming that A and B are positive, and where k is some constant.

[8]In practice one may use common logarithms, that is, log to the base 10. The relationship between the natural log and common log is: $\ln_e X = 2.3026 \log_{10} X$. By convention, ln means natural logarithm, and log means logarithm to the base 10; hence there is no need to write the subscripts e and 10 explicitly.

where $\alpha = \ln \beta_1$, this model is linear in the parameters α and β_2, linear in the logarithms of the variables Y and X, and can be estimated by OLS regression. Because of this linearity, such models are called **log-log, double-log,** or **log-linear** models.

If the assumptions of the classical linear regression model are fulfilled, the parameters of (6.5.3) can be estimated by the OLS method by letting

$$Y_i^* = \alpha + \beta_2 X_i^* + u_i \qquad \textbf{(6.5.4)}$$

where $Y_i^* = \ln Y_i$ and $X_i^* = \ln X_i$. The OLS estimators $\hat{\alpha}$ and $\hat{\beta}_2$ obtained will be best linear unbiased estimators of α and β_2, respectively.

One attractive feature of the log-log model, which has made it popular in applied work, is that the slope coefficient β_2 measures the **elasticity** of Y with respect to X, that is, the percentage change in Y for a given (small) percentage change in X.[9] Thus, if Y represents the quantity of a commodity demanded and X its unit price, β_2 measures the price elasticity of demand, a parameter of considerable economic interest. If the relationship between quantity demanded and price is as shown in Figure 6.3*a*, the double-log

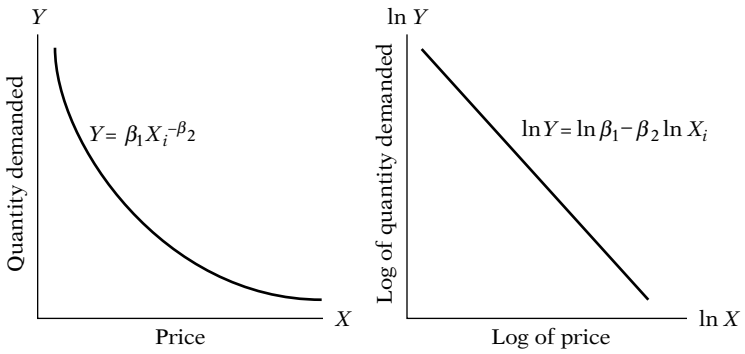

FIGURE 6.3 Constant-elasticity model.

[9]The elasticity coefficient, in calculus notation, is defined as $(dY/Y)/(dX/X) = [(dY/dX)(X/Y)]$. Readers familiar with differential calculus will readily see that β_2 is in fact the elasticity coefficient.

A technical note: The calculus-minded reader will note that $d(\ln X)/dX = 1/X$ or $d(\ln X) = dX/X$, that is, for infinitesimally small changes (note the differential operator d) the change in $\ln X$ is equal to the relative or proportional change in X. In practice, though, if the change in X is small, this relationship can be written as: change in $\ln X \doteq$ relative change in X, where $\doteq$ means approximately. Thus, for small changes,

$$(\ln X_t - \ln X_{t-1}) \doteq (X_t - X_{t-1})/X_{t-1} = \text{relative change in } X$$

Incidentally, the reader should note these terms, which will occur frequently: (1) **absolute change,** (2) **relative** or **proportional change,** and (3) **percentage change,** or **percent growth rate.** Thus, $(X_t - X_{t-1})$ represents absolute change, $(X_t - X_{t-1})/X_{t-1} = (X_t/X_{t-1} - 1)$ is relative or proportional change and $[(X_t - X_{t-1})/X_{t-1}]100$ is the percentage change, or the growth rate. X_t and X_{t-1} are, respectively, the current and previous values of the variable X.

transformation as shown in Figure 6.3*b* will then give the estimate of the price elasticity ($-\beta_2$).

Two special features of the log-linear model may be noted: The model assumes that the elasticity coefficient between Y and X, β_2, remains constant throughout (why?), hence the alternative name **constant elasticity model.**[10] In other words, as Figure 6.3*b* shows, the change in $\ln Y$ per unit change in $\ln X$ (i.e., the elasticity, β_2) remains the same no matter at which $\ln X$ we measure the elasticity. Another feature of the model is that although $\hat{\alpha}$ and $\hat{\beta}_2$ are unbiased estimates of α and β_2, β_1 (the parameter entering the original model) when estimated as $\hat{\beta}_1 = \text{antilog}\,(\hat{\alpha})$ is itself a biased estimator. In most practical problems, however, the intercept term is of secondary importance, and one need not worry about obtaining its unbiased estimate.[11]

In the two-variable model, the simplest way to decide whether the log-linear model fits the data is to plot the scattergram of $\ln Y_i$ against $\ln X_i$ and see if the scatter points lie approximately on a straight line, as in Figure 6.3*b*.

AN ILLUSTRATIVE EXAMPLE: EXPENDITURE ON DURABLE GOODS IN RELATION TO TOTAL PERSONAL CONSUMPTION EXPENDITURE

Table 6.3 presents data on total personal consumption expenditure (PCEXP), expenditure on durable goods (EXPDUR), expenditure on nondurable goods (EXPNONDUR), and expenditure on services (EXPSERVICES), all measured in 1992 billions of dollars.[12]

Suppose we wish to find the elasticity of expenditure on durable goods with respect to total personal consumption expenditure. Plotting the log of expenditure on durable goods against the log of total personal consumption expenditure, you will see that the relationship between the two variables is linear. Hence, the double-log model may be appropriate. The regression results are as follows:

$$\widehat{\ln \text{EXDUR}}_t = -9.6971 + 1.9056 \ln \text{PCEX}_t \quad \textbf{(6.5.5)}$$
$$\text{se} = (0.4341) \quad (0.0514)$$
$$t = (-22.3370)^* \quad (37.0962)^* \quad r^2 = 0.9849$$

where * indicates that the p value is extremely small.

As these results show, the elasticity of EXPDUR with respect to PCEX is about 1.90, suggesting that if total personal expenditure goes up by 1 percent, on average, the expenditure on durable goods goes up by about 1.90 percent. Thus, expenditure on durable goods is very responsive to changes in personal consumption expenditure. This is one reason why producers of durable goods keep a keen eye on changes in personal income and personal consumption expenditure. In exercises 6.17 and 6.18, the reader is asked to carry out a similar exercise for nondurable goods expenditure and expenditure on services.

(*Continued*)

[10] A constant elasticity model will give a constant total revenue change for a given percentage change in price regardless of the absolute level of price. Readers should contrast this result with the elasticity conditions implied by a simple linear demand function, $Y_i = \beta_1 + \beta_2 X_i + u_i$. However, a simple linear function gives a constant quantity change per unit change in price. Contrast this with what the log-linear model implies for a given dollar change in price.

[11] Concerning the nature of the bias and what can be done about it, see Arthur S. Goldberger, *Topics in Regression Analysis*, Macmillan, New York, 1978, p. 120.

[12] Durable goods include motor vehicles and parts, furniture, and household equipment; nondurable goods include food, clothing, gasoline and oil, fuel oil and coal; and services include housing, electricity and gas, transportation, and medical care.

AN ILLUSTRATIVE EXAMPLE: . . . (*Continued*)

TABLE 6.3
TOTAL PERSONAL EXPENDITURE AND CATEGORIES

Observation	EXPSERVICES	EXPDUR	EXPNONDUR	PCEXP
1993-I	2445.3	504.0	1337.5	4286.8
1993-II	2455.9	519.3	1347.8	4322.8
1993-III	2480.0	529.9	1356.8	4366.6
1993-IV	2494.4	542.1	1361.8	4398.0
1994-I	2510.9	550.7	1378.4	4439.4
1994-II	2531.4	558.8	1385.5	4472.2
1994-III	2543.8	561.7	1393.2	4498.2
1994-IV	2555.9	576.6	1402.5	4534.1
1995-I	2570.4	575.2	1410.4	4555.3
1995-II	2594.8	583.5	1415.9	4593.6
1995-III	2610.3	595.3	1418.5	4623.4
1995-IV	2622.9	602.4	1425.6	4650.0
1996-I	2648.5	611.0	1433.5	4692.1
1996-II	2668.4	629.5	1450.4	4746.6
1996-III	2688.1	626.5	1454.7	4768.3
1996-IV	2701.7	637.5	1465.1	4802.6
1997-I	2722.1	656.3	1477.9	4853.4
1997-II	2743.6	653.8	1477.1	4872.7
1997-III	2775.4	679.6	1495.7	4947.0
1997-IV	2804.8	648.8	1494.3	4981.0
1998-I	2829.3	710.3	1521.2	5055.1
1998-II	2866.8	729.4	1540.9	5130.2
1998-III	2904.8	733.7	1549.1	5181.8

Note: EXPSERVICES = expenditure on services, billions of 1992 dollars.
EXPDUR = expenditure on durable goods, billions of 1992 dollars.
EXPNONDUR = expenditure on nondurable goods, billions of 1992 dollars.
PCEXP = total personal consumption expenditure, billions of 1992 dollars.
Source: Economic Report of the President, 1999, Table B-17, p. 347.

6.6 SEMILOG MODELS: LOG–LIN AND LIN–LOG MODELS

How to Measure the Growth Rate: The Log–Lin Model

Economists, businesspeople, and governments are often interested in finding out the rate of growth of certain economic variables, such as population, GNP, money supply, employment, productivity, and trade deficit.

Suppose we want to find out the growth rate of personal consumption expenditure on services for the data given in Table 6.3. Let Y_t denote real expenditure on services at time t and Y_0 the initial value of the expenditure on services (i.e., the value at the end of 1992-IV). You may recall the following well-known compound interest formula from your introductory course in economics.

$$Y_t = Y_0(1 + r)^t \tag{6.6.1}$$

where r is the compound (i.e., over time) rate of growth of Y. Taking the natural logarithm of (6.6.1), we can write

$$\ln Y_t = \ln Y_0 + t \ln (1 + r) \tag{6.6.2}$$

Now letting

$$\beta_1 = \ln Y_0 \tag{6.6.3}$$

$$\beta_2 = \ln (1 + r) \tag{6.6.4}$$

we can write (6.6.2) as

$$\ln Y_t = \beta_1 + \beta_2 t \tag{6.6.5}$$

Adding the disturbance term to (6.6.5), we obtain[13]

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \tag{6.6.6}$$

This model is like any other linear regression model in that the parameters β_1 and β_2 are linear. The only difference is that the regressand is the logarithm of Y and the regressor is "time," which will take values of 1, 2, 3, etc.

Models like (6.6.6) are called **semilog models** because only one variable (in this case the regressand) appears in the logarithmic form. For descriptive purposes a model in which the regressand is logarithmic will be called a **log-lin model.** Later we will consider a model in which the regressand is linear but the regressor(s) are logarithmic and call it a **lin-log model.**

Before we present the regression results, let us examine the properties of model (6.6.5). In this model *the slope coefficient measures the constant proportional or relative change in Y for a given absolute change in the value of the regressor* (in this case the variable t), that is,[14]

$$\beta_2 = \frac{\text{relative change in regressand}}{\text{absolute change in regressor}} \tag{6.6.7}$$

If we multiply the relative change in Y by 100, (6.6.7) will then give the percentage change, or the *growth rate*, in Y for an absolute change in X, the regressor. That is, 100 times β_2 gives the growth rate in Y; 100 times β_2 is

[13]We add the error term because the compound interest formula will not hold exactly. Why we add the error after the logarithmic transformation is explained in Sec. 6.8.

[14]Using differential calculus one can show that $\beta_2 = d(\ln Y)/dX = (1/Y)(dY/dX) = (dY/Y)/dX$, which is nothing but (6.6.7). For small changes in Y and X this relation may be approximated by

$$\frac{(Y_t - Y_{t-1})/Y_{t-1}}{(X_t - X_{t-1})}$$

Note: Here $X = t$.

known in the literature as the **semielasticity** of Y with respect to X. (Question: To get the elasticity, what will we have to do?)

AN ILLUSTRATIVE EXAMPLE: THE RATE OF GROWTH EXPENDITURE ON SERVICES

To illustrate the growth model (6.6.6), consider the data on expenditure on services given in Table 6.3. The regression results are as follows:

$$\widehat{\ln \text{EXS}}_t = \underset{\text{se} =}{7.7890} + \underset{(0.0023)}{0.00743}\underset{(0.00017)}{t} \qquad \textbf{(6.6.8)}$$
$$t = (3387.619)^* \quad (44.2826)^* \quad r^2 = 0.9894$$

Note: EXS stands for expenditure on services and * denotes that the *p* value is extremely small.

The interpretation of Eq. (6.6.8) is that over the quarterly period 1993:1 to 1998:3, expenditure on services increased at the (quarterly) rate of 0.743 percent. Roughly, this is equal to an annual growth rate of 2.97 percent. Since 7.7890 = log of EXS at the beginning of the study period, by taking its antilog we obtain 2413.90 (billion dollars) as the beginning value of EXS (i.e., the value at the end of the fourth quarter of 1992). The regression line obtained in Eq. (6.6.8) is sketched in Figure 6.4.

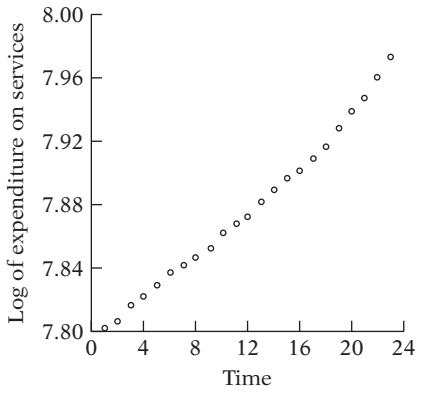

FIGURE 6.4

Instantaneous versus Compound Rate of Growth. The coefficient of the trend variable in the growth model (6.6.6), β_2, gives the **instantaneous** (at a point in time) rate of growth and not the **compound** (over a period of time) rate of growth. But the latter can be easily found from (6.6.4) by taking the antilog of the estimated β_2 and subtracting 1 from it and multiplying the difference by 100. Thus, for our illustrative example, the estimated slope coefficient is 0.00743. Therefore, [antilog(0.00743) − 1] = 0.00746 or 0.746 percent. Thus, in the illustrative example, the *compound rate of growth* on expenditure on services was about 0.746 percent per quarter, which is slightly higher than the instantaneous growth rate of 0.743 percent. This is of course due to the compounding effect.

Linear Trend Model. Instead of estimating model (6.6.6), researchers sometimes estimate the following model:

$$Y_t = \beta_1 + \beta_2 t + u_t \qquad \textbf{(6.6.9)}$$

That is, instead of regressing the log of Y on time, they regress Y on time, where Y is the regressand under consideration. Such a model is called a **linear trend model** and the time variable t is known as the *trend variable*. If the slope coefficient in (6.6.9) is positive, there is an **upward trend** in Y, whereas if it is negative, there is a **downward trend** in Y.

For the expenditure on services data that we considered earlier, the results of fitting the linear trend model (6.6.9) are as follows:

$$\begin{aligned} \widehat{\text{EXS}}_t &= 2405.848 \quad + \quad 19.6920t \\ t &= (322.9855) \quad (36.2479) \qquad r^2 = 0.9843 \end{aligned} \tag{6.6.10}$$

In contrast to Eq. (6.6.8), the interpretation of Eq. (6.6.10) is as follows: Over the quarterly period 1993-I to 1998-III, on average, expenditure on services increased at the absolute (*note:* not relative) rate of about 20 billion dollars per quarter. That is, there was an upward trend in the expenditure on services.

The choice between the growth rate model (6.6.8) and the linear trend model (6.6.10) will depend upon whether one is interested in the relative or absolute change in the expenditure on services, although for comparative purposes it is the relative change that is generally more relevant. In passing, *observe that we cannot compare the r^2 values of models (6.6.8) and (6.6.10) because the regressands in the two models are different*. We will show in Chapter 7 how one compares the R^2's of models like (6.6.8) and (6.6.10).

The Lin–Log Model

Unlike the growth model just discussed, in which we were interested in finding the percent growth in Y for an absolute change in X, suppose we now want to find the absolute change in Y for a percent change in X. A model that can accomplish this purpose can be written as:

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i \tag{6.6.11}$$

For descriptive purposes we call such a model a **lin–log model.**

Let us interpret the slope coefficient β_2.[15] As usual,

$$\begin{aligned} \beta_2 &= \frac{\text{change in } Y}{\text{change in } \ln X} \\ &= \frac{\text{change in } Y}{\text{relative change in } X} \end{aligned}$$

The second step follows from the fact that *a change in the log of a number is a relative change.*

[15]Again, using differential calculus, we have

$$\frac{dY}{dX} = \beta_2\left(\frac{1}{X}\right)$$

Therefore,

$$\beta_2 = \frac{dY}{\frac{dX}{X}} = (6.6.12)$$

Symbolically, we have

$$\beta_2 = \frac{\Delta Y}{\Delta X/X} \tag{6.6.12}$$

where, as usual, Δ denotes a small change. Equation (6.6.12) can be written, equivalently, as

$$\Delta Y = \beta_2(\Delta X/X) \tag{6.6.13}$$

This equation states that the absolute change in Y ($= \Delta Y$) is equal to slope times the relative change in X. If the latter is multiplied by 100, then (6.6.13) gives the absolute change in Y for a percentage change in X. Thus, if $(\Delta X/X)$ changes by 0.01 unit (or 1 percent), the absolute change in Y is $0.01(\beta_2)$; if in an application one finds that $\beta_2 = 500$, the absolute change in Y is $(0.01)(500) = 5.0$. Therefore, when regression (6.6.11) is estimated by OLS, do not forget to multiply the value of the estimated slope coefficient by 0.01, or, what amounts to the same thing, divide it by 100. *If you do not keep this in mind, your interpretation in an application will be highly misleading.*

The practical question is: When is a lin–log model like (6.6.11) useful? An interesting application has been found in the so-called **Engel expenditure** models, named after the German statistician Ernst Engel, 1821–1896. (See exercise 6.10.) Engel postulated that "the total expenditure that is devoted to food tends to increase in arithmetic progression as total expenditure increases in geometric progression."[16]

AN ILLUSTRATIVE EXAMPLE

As an illustration of the lin–log model, let us revisit our example on food expenditure in India, Example 3.2. There we fitted a linear-in-variables model as a first approximation. But if we plot the data we obtain the plot in Figure 6.5. As this figure suggests, food expenditure increases more slowly as total expenditure increases, perhaps giving credence to Engel's law. The results of fitting the lin–log model to the data are as follows:

$$\widehat{\text{FoodExp}}_i = -1283.912 + 257.2700 \ln \text{TotalExp}_i$$
$$t = \quad (-4.3848)^* \quad (5.6625)^* \quad r^2 = 0.3769 \tag{6.6.14}$$

Note: * denotes an extremely small *p value.*

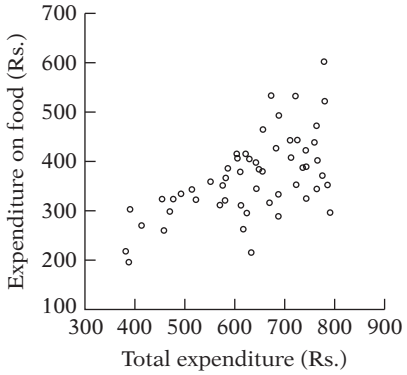

FIGURE 6.5

(*Continued*)

[16]See Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries,* Routledge, London, 1998, p. 158. This quote is attributed to H. Working, "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association,* vol. 38, 1943, pp. 43–56.

AN ILLUSTRATIVE EXAMPLE (*Continued*)

Interpreted in the manner described earlier, the slope coefficient of about 257 means that an increase in the total food expenditure of 1 percent, on average, leads to about 2.57 rupees increase in the expenditure on food of the 55 families included in the sample. (*Note:* We have divided the estimated slope coefficient by 100.)

Before proceeding further, note that if you want to compute the elasticity coefficient for the log–lin or lin–log models, you can do so from the definition of the elasticity coefficient given before, namely,

$$\text{Elasticity} = \frac{dY}{dX}\frac{X}{Y}$$

As a matter of fact, once the functional form of a model is known, one can compute elasticities by applying the preceding definition. (Table 6.6, given later, summarizes the elasticity coefficients for the various models.)

6.7 RECIPROCAL MODELS

Models of the following type are known as **reciprocal** models.

$$Y_i = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i \tag{6.7.1}$$

Although this model is nonlinear in the variable X because it enters inversely or reciprocally, the model is linear in β_1 and β_2 and is therefore a linear regression model.[17]

This model has these features: As X increases indefinitely, the term $\beta_2(1/X)$ approaches zero (*note:* β_2 is a constant) and Y approaches the limiting or *asymptotic* value β_1. Therefore, models like (6.7.1) have built in them an **asymptote** or limit value that the dependent variable will take when the value of the X variable increases indefinitely.[18]

Some likely shapes of the curve corresponding to (6.7.1) are shown in Figure 6.6. As an illustration of Figure 6.6*a*, consider the data given in Table 6.4. These are cross-sectional data for 64 countries on child mortality and a few other variables. For now, concentrate on the variables, child mortality (CM) and per capita GNP, which are plotted in Figure 6.7.

As you can see, this figure resembles Figure 6.6*a*: As per capita GNP increases, one would expect child mortality to decrease because people can afford to spend more on health care, assuming all other factors remain constant. But the relationship is not a straight line one: As per capita GNP increases, initially there is dramatic drop in CM but the drop tapers off as per capita GNP continues to increase.

[17]If we let $X_i^* = (1/X_i)$, then (6.7.1) is linear in the parameters as well as the variables Y_i and X_i^*.

[18]The slope of (6.7.1) is: $dY/dX = -\beta_2(1/X^2)$, implying that if β_2 is positive, the slope is negative throughout, and if β_2 is negative, the slope is positive throughout. See Figures 6.6*a* and 6.6*c*, respectively.

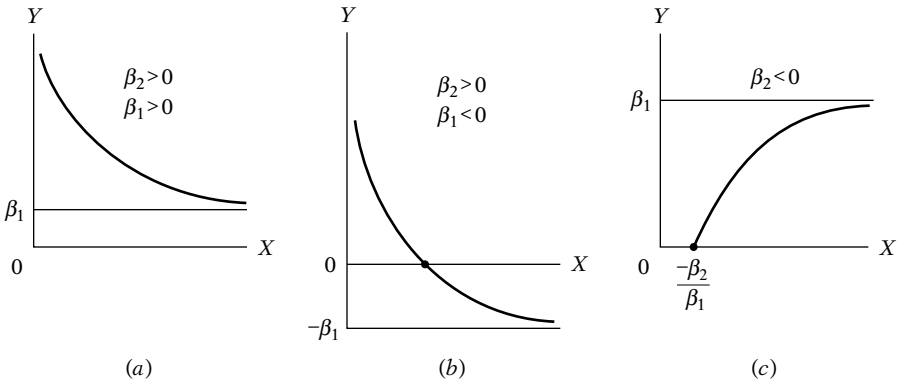

FIGURE 6.6 The reciprocal model: $Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$.

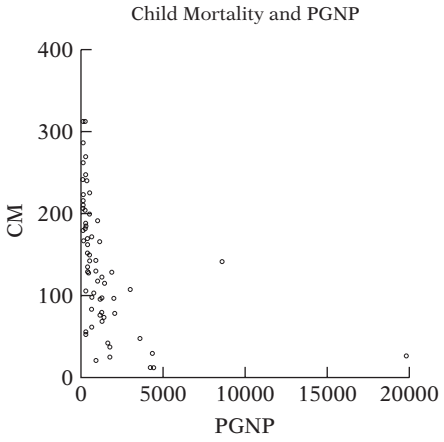

FIGURE 6.7 Relationship between child mortality and per capita GNP in 66 countries.

If we try to fit the reciprocal model (6.7.1), we obtain the following regression results:

$$\begin{aligned} \widehat{\text{CM}}_i &= 81.79436 + 27{,}273.17\left(\frac{1}{\text{PGNP}_i}\right) \\ \text{se} &= (10.8321) \qquad (3759.999) \\ t &= (7.5511) \qquad (7.2535) \qquad r^2 = 0.4590 \end{aligned} \tag{6.7.2}$$

As per capita GNP increases indefinitely, child mortality approaches its asymptotic value of about 82 deaths per thousand. As explained in footnote 18, the positive value of the coefficient of $(1/\text{PGNP}_t)$ implies that the rate of change of CM with respect to PGNP is negative.

One of the important applications of Figure 6.6*b* is the celebrated Phillips curve of macroeconomics. Using the data on percent rate of change of money wages (Y) and the unemployment rate (X) for the United Kingdom

TABLE 6.4 FERTILITY AND OTHER DATA FOR 64 COUNTRIES

Observation	CM	FLFP	PGNP	TFR	Observation	CM	FLFP	PGNP	TFR
1	128	37	1870	6.66	33	142	50	8640	7.17
2	204	22	130	6.15	34	104	62	350	6.60
3	202	16	310	7.00	35	287	31	230	7.00
4	197	65	570	6.25	36	41	66	1620	3.91
5	96	76	2050	3.81	37	312	11	190	6.70
6	209	26	200	6.44	38	77	88	2090	4.20
7	170	45	670	6.19	39	142	22	900	5.43
8	240	29	300	5.89	40	262	22	230	6.50
9	241	11	120	5.89	41	215	12	140	6.25
10	55	55	290	2.36	42	246	9	330	7.10
11	75	87	1180	3.93	43	191	31	1010	7.10
12	129	55	900	5.99	44	182	19	300	7.00
13	24	93	1730	3.50	45	37	88	1730	3.46
14	165	31	1150	7.41	46	103	35	780	5.66
15	94	77	1160	4.21	47	67	85	1300	4.82
16	96	80	1270	5.00	48	143	78	930	5.00
17	148	30	580	5.27	49	83	85	690	4.74
18	98	69	660	5.21	50	223	33	200	8.49
19	161	43	420	6.50	51	240	19	450	6.50
20	118	47	1080	6.12	52	312	21	280	6.50
21	269	17	290	6.19	53	12	79	4430	1.69
22	189	35	270	5.05	54	52	83	270	3.25
23	126	58	560	6.16	55	79	43	1340	7.17
24	12	81	4240	1.80	56	61	88	670	3.52
25	167	29	240	4.75	57	168	28	410	6.09
26	135	65	430	4.10	58	28	95	4370	2.86
27	107	87	3020	6.66	59	121	41	1310	4.88
28	72	63	1420	7.28	60	115	62	1470	3.89
29	128	49	420	8.12	61	186	45	300	6.90
30	27	63	19830	5.23	62	47	85	3630	4.10
31	152	84	420	5.79	63	178	45	220	6.09
32	224	23	530	6.50	64	142	67	560	7.20

Note: CM = Child mortality, the number of deaths of children under age 5 in a year per 1000 live births.
FLFP = Female literacy rate, percent.
PGNP = per capita GNP in 1980.
TFR = total fertility rate, 1980–1985, the average number of children born to a woman, using age-specific fertility rates for a given year.
Source: Chandan Mukherjee, Howard White, and Marc Whyte, *Econometrics and Data Analysis for Developing Countries,* Routledge, London, 1998, p. 456.

for the period 1861–1957, Phillips obtained a curve whose general shape resembles Figure 6.6*b* (Figure 6.8).[19]

As Figure 6.8 shows, there is an asymmetry in the response of wage changes to the level of the unemployment rate: Wages rise faster for a unit change in unemployment if the unemployment rate is below U^n, which is

[19]A. W. Phillips, "The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861–1957," *Economica,* November 1958, vol. 15, pp. 283–299. Note that the original curve did not cross the unemployment rate axis, but Fig. 6.8 represents a later version of the curve.

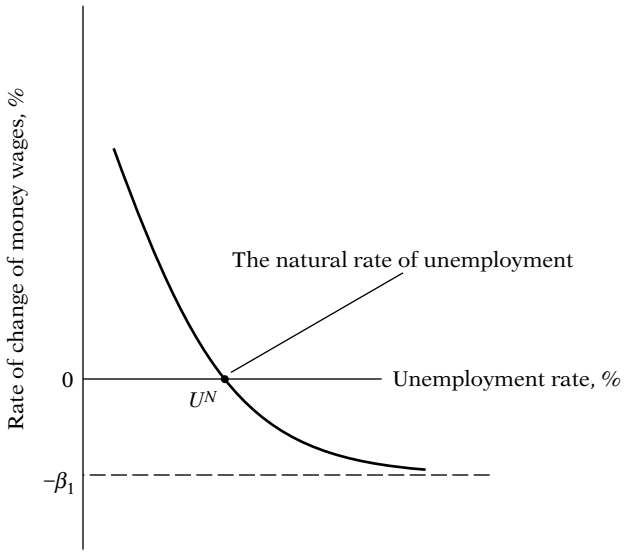

FIGURE 6.8 The Phillips curve.

called the *natural rate of unemployment* by economists [defined as the rate of unemployment required to keep (wage) inflation constant], and then they fall for an equivalent change when the unemployment rate is above the natural rate, β_1, indicating the asymptotic floor for wage change. This particular feature of the Phillips curve may be due to institutional factors, such as union bargaining power, minimum wages, unemployment compensation, etc.

Since the publication of Phillips' article, there has been very extensive research on the Phillips curve at the theoretical as well as empirical levels. Space does not permit us to go into the details of the controversy surrounding the Phillips curve. The Phillips curve itself has gone through several incarnations. A comparatively recent formulation is provided by Olivier Blanchard.[20] If we let π_t denote the inflation rate at time t, which is defined as the percentage change in the price level as measured by a representative price index, such as the Consumer Price Index (CPI), and UN_t denote the unemployment rate at time t, then a modern version of the Phillips curve can be expressed in the following format:

$$\pi_t - \pi_t^e = \beta_2(\text{UN}_t - U^n) + u_t \qquad \textbf{(6.7.3)}$$

where π_t = actual inflation rate at time t
π_t^e = expected inflation rate at time t, the expectation being formed in year $(t-1)$

[20]See Olivier Blanchard, *Macroeconomics,* Prentice Hall, Englewood Cliffs, N.J., 1997, Chap. 17.

UN_t = actual unemployment rate prevailing at time t
U^n = natural rate of unemployment at time t
u_t = stochastic error term[21]

Since π_t^e is not directly observable, as a starting point one can make the simplifying assumption that $\pi_t^e = \pi_{t-1}$; that is, the inflation expected this year is the inflation rate that prevailed in the last year; of course, more complicated assumptions about expectations formation can be made, and we will discuss this topic in Chapter 17, on distributed lag models.

Substituting this assumption into (6.7.3) and writing the regression model in the standard form, we obtain the following estimating equation:

$$\pi_t - \pi_{t-1} = \beta_1 + \beta_2 \text{UN}_t + u_t \tag{6.7.4}$$

where $\beta_1 = -\beta_2 U^n$. Equation (6.7.4) states that the change in the inflation rate between two time periods is linearly related to the current unemployment rate. A priori, β_2 is expected to be negative (why?) and β_1 is expected to be positive (this figures, since β_2 is negative and U^n is positive).

Incidentally, the Phillips relationship given in (6.7.3) is known in the literature as the **modified Phillips curve,** or the **expectations-augmented Phillips curve** (to indicate that π_{t-1} stands for expected inflation), or the **acceleratonist Phillips curve** (to suggest that a low unemployment rate leads to an increase in the inflation rate and hence an *acceleration* of the price level).

As an illustration of the modified Phillips curve, we present in Table 6.5 data on inflation as measured by year-to-year percentage in the Consumer Price Index (CPIflation) and the unemployment rate for the period 1960–1998. The unemployment rate represents the civilian unemployment rate. From these data we obtained the change in the inflation rate $(\pi_t - \pi_{t-1})$ and plotted it against the civilian unemployment rate; we are using the CPI as a measure of inflation. The resulting graph appears in Figure 6.9.

As expected, the relation between the change in inflation rate and the unemployment rate is negative—a low unemployment rate leads to an increase in the inflation rate and therefore an acceleration of the price level, hence the name accelerationist Phillips curve.

Looking at Figure 6.9, it is not obvious whether a linear (straight line) regression model or a reciprocal model fits the data; there may be a curvilinear relationship between the two variables. We present below regressions based on both the models. However, keep in mind that for the reciprocal model the intercept term is expected to be negative and the slope positive, as noted in footnote 18.

$$\text{Linear model:} \quad \widehat{(\pi_t - \pi_{t-1})} = 4.1781 - 0.6895\ \text{UN}_t \qquad t = (3.9521) \quad (-4.0692) \qquad r^2 = 0.3150 \tag{6.7.5}$$

[21]Economists believe this error term represents some kind of supply shock, such as the OPEC oil embargoes of 1973 and 1979.

TABLE 6.5 INFLATION RATE AND UNEMPLOYMENT RATE, UNITED STATES, 1960–1998

Observation	INFLRATE	UNRATE	Observation	INFLRATE	UNRATE
1960	1.7	5.5	1980	13.5	7.1
1961	1.0	6.7	1981	10.3	7.6
1962	1.0	5.5	1982	6.2	9.7
1963	1.3	5.7	1983	3.2	9.6
1964	1.3	5.2	1984	4.3	7.5
1965	1.6	4.5	1985	3.6	7.2
1966	2.9	3.8	1986	1.9	7.0
1967	3.1	3.8	1987	3.6	6.2
1968	4.2	3.6	1988	4.1	5.5
1969	5.5	3.5	1989	4.8	5.3
1970	5.7	4.9	1990	5.4	5.6
1971	4.4	5.9	1991	4.2	6.8
1972	3.2	5.6	1992	3.0	7.5
1973	6.2	4.9	1993	3.0	6.9
1974	11.0	5.6	1994	2.6	6.1
1975	9.1	8.5	1995	2.8	5.6
1976	5.8	7.7	1996	3.0	5.4
1977	6.5	7.1	1997	2.3	4.9
1978	7.6	6.1	1998	1.6	4.5
1979	11.3	5.8			

Note: The inflation rate is the percent year-to-year change in CPI. The unemployment rate is the civilian unemployment rate.

Source: Economic Report of the President, 1999, Table B-63, p. 399, for CPI changes and Table B-42, p. 376, for the unemployment rate.

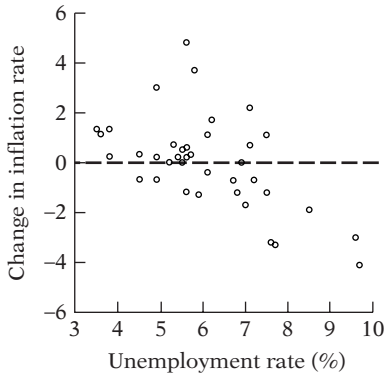

FIGURE 6.9 The modified Phillips curve.

Reciprocal model:

$$\widehat{(\pi_t - \pi_{t-1})} = -3.2514 + 18.5508\left(\frac{1}{\text{UN}_t}\right) \qquad \textbf{(6.7.6)}$$

$$t = (-2.9715) \quad (3.0625) \qquad r^2 = 0.2067$$

All the estimated coefficients in both the models are *individually* statistically significant, all the p values being lower than the 0.005 level.

Model (6.7.5) shows that if the unemployment rate goes down by 1 percentage point, on average, the change in the inflation rate goes up by about 0.7 percentage points, and vice versa. Model (6.7.6) shows that even if the unemployment rate increases indefinitely, the most the change in the inflation rate will go down will be about 3.25 percentage points. Incidentally, from Eq. (6.7.5), we can compute the underlying natural rate of unemployment as:

$$U^n = \frac{\hat{\beta}_1}{-\hat{\beta}_2} = \frac{4.1781}{0.6895} = 6.0596 \qquad \textbf{(6.7.7)}$$

That is, the natural rate of unemployment is about 6.06%. Economists put the natural rate between 5 to 6%, although in the recent past in the United States the actual rate has been much below this rate.

Log Hyperbola or Logarithmic Reciprocal Model

We conclude our discussion of reciprocal models by considering the logarithmic reciprocal model, which takes the following form:

$$\ln Y_i = \beta_1 - \beta_2 \left(\frac{1}{X_i}\right) + u_i \qquad \textbf{(6.7.8)}$$

Its shape is as depicted in Figure 6.10. As this figure shows, initially Y increases at an increasing rate (i.e., the curve is initially convex) and then it increases at a decreasing rate (i.e., the curve becomes concave).[22] Such a

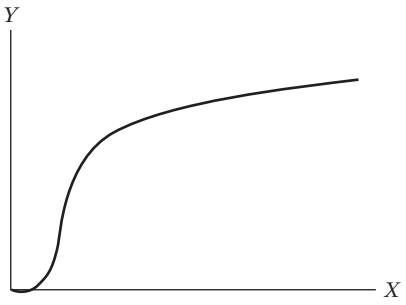

FIGURE 6.10 The log reciprocal model.

[22]From calculus, it can be shown that

$$\frac{d}{dX}(\ln Y) = -\beta_2\left(-\frac{1}{X^2}\right) = \beta_2\left(\frac{1}{X^2}\right)$$

But

$$\frac{d}{dX}(\ln Y) = \frac{1}{Y}\frac{dY}{dX}$$

Making this substitution, we obtain

$$\frac{dY}{dX} = \beta_2 \frac{Y}{X^2}$$

which is the slope of Y with respect to X.

model may therefore be appropriate to model a short-run production function. Recall from microeconomics that if labor and capital are the inputs in a production function and if we keep the capital input constant but increase the labor input, the short-run output–labor relationship will resemble Figure 6.10. (See Example 7.4, Chapter 7.)

6.8 CHOICE OF FUNCTIONAL FORM

In this chapter we discussed several functional forms an empirical model can assume, even within the confines of the linear-in-parameter regression models. The choice of a particular functional form may be comparatively easy in the two-variable case, because we can plot the variables and get some rough idea about the appropriate model. The choice becomes much harder when we consider the multiple regression model involving more than one regressor, as we will discover when we discuss this topic in the next two chapters. There is no denying that a great deal of skill and experience are required in choosing an appropriate model for empirical estimation. But some guidelines can be offered:

1. The underlying theory (e.g., the Phillips curve) may suggest a particular functional form.

2. It is good practice to find out the rate of change (i.e., the slope) of the regressand with respect to the regressor as well as to find out the elasticity of the regressand with respect to the regressor. For the various models considered in this chapter, we provide the necessary formulas for the slope and elasticity coefficients of the various models in Table 6.6. The knowledge of these formulas will help us to compare the various models.

TABLE 6.6

Model	Equation	Slope $\left(=\frac{dY}{dX}\right)$	Elasticity $\left(=\frac{dY}{dX}\frac{X}{Y}\right)$
Linear	$Y = \beta_1 + \beta_2 X$	β_2	$\beta_2\left(\frac{X}{Y}\right)^*$
Log–linear	$\ln Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{Y}{X}\right)$	β_2
Log–lin	$\ln Y = \beta_1 + \beta_2 X$	$\beta_2 (Y)$	$\beta_2 (X)^*$
Lin–log	$Y = \beta_1 + \beta_2 \ln X$	$\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{1}{Y}\right)^*$
Reciprocal	$Y = \beta_1 + \beta_2\left(\frac{1}{X}\right)$	$-\beta_2\left(\frac{1}{X^2}\right)$	$-\beta_2\left(\frac{1}{XY}\right)^*$
Log reciprocal	$\ln Y = \beta_1 - \beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{Y}{X^2}\right)$	$\beta_2\left(\frac{1}{X}\right)^*$

Note: * indicates that the elasticity is variable, depending on the value taken by X or Y or both. When no X and Y values are specified, in practice, very often these elasticities are measured at the mean values of these variables, namely, $\bar{X}$ and $\bar{Y}$.

3. The coefficients of the model chosen should satisfy certain a priori expectations. For example, if we are considering the demand for automobiles as a function of price and some other variables, we should expect a negative coefficient for the price variable.

4. Sometime more than one model may fit a given set of data reasonably well. In the modified Phillips curve, we fitted both a linear and a reciprocal model to the same data. In both cases the coefficients were in line with prior expectations and they were all statistically significant. One major difference was that the r^2 value of the linear model was larger than that of the reciprocal model. One may therefore give a slight edge to the linear model over the reciprocal model. *But make sure that in comparing two r^2 values the dependent variable, or the regressand, of the two models is the same; the regressor(s) can take any form.* We will explain the reason for this in the next chapter.

5. In general *one should not overemphasize* the r^2 measure in the sense that the higher the r^2 the better the model. As we will discuss in the next chapter, r^2 increases as we add more regressors to the model. What is of greater importance is the theoretical underpinning of the chosen model, the signs of the estimated coefficients and their statistical significance. If a model is good on these criteria, a model with a lower r^2 may be quite acceptable. We will revisit this important topic in greater depth in Chapter 13.

*6.9 A NOTE ON THE NATURE OF THE STOCHASTIC ERROR TERM: ADDITIVE VERSUS MULTIPLICATIVE STOCHASTIC ERROR TERM

Consider the following regression model, which is the same as (6.5.1) but without the error term:

$$Y_i = \beta_1 X^{\beta_2} \tag{6.9.1}$$

For estimation purposes, we can express this model in three different forms:

$$Y_i = \beta_1 X_i^{\beta_2} u_i \tag{6.9.2}$$

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \tag{6.9.3}$$

$$Y_i = \beta_1 X_i^{\beta_2} + u_i \tag{6.9.4}$$

Taking the logarithms on both sides of these equations, we obtain

$$\ln Y_i = \alpha + \beta_2 \ln X_i + \ln\ u_i \tag{6.9.2a}$$

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \tag{6.9.3a}$$

$$\ln Y_i = \ln\left(\beta_1 X_i^{\beta_2} + u_i\right) \tag{6.9.4a}$$

where $\alpha = \ln \beta_1$.

*Optional

Models like (6.9.2) are *intrinsically linear (in-parameter)* regression models in the sense that by suitable (log) transformation the models can be made linear in the parameters α and β_2. (*Note:* These models are nonlinear in β_1.) But model (6.9.4) is *intrinsically nonlinear-in-parameter.* There is no simple way to take the log of (6.9.4) because $\ln(A + B) \neq \ln A + \ln B$.

Although (6.9.2) and (6.9.3) are linear regression models and can be estimated by OLS or ML, we have to be careful about the properties of the stochastic error term that enters these models. Remember that the BLUE property of OLS requires that u_i has zero mean value, constant variance, and zero autocorrelation. For hypothesis testing, we further assume that u_i follows the normal distribution with mean and variance values just discussed. In short, we have assumed that $u_i \sim N(0, \sigma^2)$.

Now consider model (6.9.2). Its statistical counterpart is given in (6.9.2*a*). To use the classical normal linear regression model (CNLRM), we have to assume that

$$\ln u_i \sim N(0, \sigma^2) \tag{6.9.5}$$

Therefore, when we run the regression (6.9.2*a*), we will have to apply the normality tests discussed in Chapter 5 to the residuals obtained from this regression. Incidentally, note that if $\ln u_i$ follows the normal distribution with zero mean and constant variance, then statistical theory shows that u_i in (6.9.2) must follow the **log-normal distribution** with mean $e^{\sigma^2/2}$ and variance $e^{\sigma^2}(e^{\sigma^2} - 1)$.

As the preceding analysis shows, one has to pay very careful attention to the error term in transforming a model for regression analysis. As for (6.9.4), this model is a *nonlinear-in-parameter* regression model and will have to be solved by some iterative computer routine. Model (6.9.3) should not pose any problems for estimation.

To sum up, pay very careful attention to the disturbance term when you transform a model for regression analysis. Otherwise, a blind application of OLS to the transformed model will not produce a model with desirable statistical properties.

6.10 SUMMARY AND CONCLUSIONS

This chapter introduced several of the finer points of the classical linear regression model (CLRM).

1. Sometimes a regression model may not contain an explicit intercept term. Such models are known as **regression through the origin.** Although the algebra of estimating such models is simple, one should use such models with caution. In such models the sum of the residuals $\sum \hat{u}_i$ is nonzero; additionally, the conventionally computed r^2 may not be meaningful. Unless

there is a strong theoretical reason, it is better to introduce the intercept in the model explicitly.

2. The units and scale in which the regressand and the regressor(s) are expressed are very important because the interpretation of regression coefficients critically depends on them. In empirical research the researcher should not only quote the sources of data but also state explicitly how the variables are measured.

3. Just as important is the functional form of the relationship between the regressand and the regressor(s). Some of the important functional forms discussed in this chapter are (*a*) the log-linear or constant elasticity model, (*b*) semilog regression models, and (*c*) reciprocal models.

4. In the log-linear model both the regressand and the regressor(s) are expressed in the logarithmic form. The regression coefficient attached to the log of a regressor is interpreted as the elasticity of the regressand with respect to the regressor.

5. In the semilog model either the regressand or the regressor(s) are in the log form. In the semilog model where the regressand is logarithmic and the regressor X is time, the estimated slope coefficient (multiplied by 100) measures the (instantaneous) rate of growth of the regressand. Such models are often used to measure the growth rate of many economic phenomena. In the semilog model if the regressor is logarithmic, its coefficient measures the absolute rate of change in the regressand for a given percent change in the value of the regressor.

6. In the reciprocal models, either the regressand or the regressor is expressed in reciprocal, or inverse, form to capture nonlinear relationships between economic variables, as in the celebrated Phillips curve.

7. In choosing the various functional forms, great attention should be paid to the stochastic disturbance term u_i. As noted in Chapter 5, the CLRM explicitly assumes that the disturbance term has zero mean value and constant (homoscedastic) variance and that it is uncorrelated with the regressor(s). It is under these assumptions that the OLS estimators are BLUE. Further, under the CNLRM, the OLS estimators are also normally distributed. One should therefore find out if these assumptions hold in the functional form chosen for empirical analysis. After the regression is run, the researcher should apply diagnostic tests, such as the normality test, discussed in Chapter 5. This point cannot be overemphasized, for the classical tests of hypothesis, such as the t, F, and χ^2, rest on the assumption that the disturbances are normally distributed. This is especially critical if the sample size is small.

8. Although the discussion so far has been confined to two-variable regression models, the subsequent chapters will show that in many cases the extension to multiple regression models simply involves more algebra without necessarily introducing more fundamental concepts. That is why it is so very important that the reader have a firm grasp of the two-variable regression model.

EXERCISES

Questions

6.1. Consider the regression model

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

where $y_i = (Y_i - \bar{Y})$ and $x_i = (X_i - \bar{X})$. In this case, the regression line must pass through the origin. True or false? Show your calculations.

6.2. The following regression results were based on monthly data over the period January 1978 to December 1987:

$$\begin{aligned}
\hat{Y}_t &= \ 0.00681 \quad + \ 0.75815X_t \\
\text{se} &= (0.02596) \quad (0.27009) \\
t &= (0.26229) \quad (2.80700) \\
p \text{ value} &= (0.7984) \quad (0.0186) \qquad r^2 = 0.4406 \\
\hat{Y}_t &= \ 0.76214X_t \\
\text{se} &= (0.265799) \\
t &= (2.95408) \\
p \text{ value} &= (0.0131) \qquad\qquad r^2 = 0.43684
\end{aligned}$$

where Y = monthly rate of return on Texaco common stock, %, and X = monthly market rate of return,%.*

a. What is the difference between the two regression models?
b. Given the preceding results, would you retain the intercept term in the first model? Why or why not?
c. How would you interpret the slope coefficients in the two models?
d. What is the theory underlying the two models?
e. Can you compare the r^2 terms of the two models? Why or why not?
f. The Jarque–Bera normality statistic for the first model in this problem is 1.1167 and for the second model it is 1.1170. What conclusions can you draw from these statistics?
g. The t value of the slope coefficient in the zero intercept model is about 2.95, whereas that with the intercept present is about 2.81. Can you rationalize this result?

6.3. Consider the following regression model:

$$\frac{1}{Y_i} = \beta_1 + \beta_2 \left(\frac{1}{X_i}\right) + u_i$$

Note: Neither Y nor X assumes zero value.
a. Is this a linear regression model?
b. How would you estimate this model?

*The underlying data were obtained from the data diskette included in Ernst R. Berndt, *The Practice of Econometrics: Classic and Contemporary,* Addison-Wesley, Reading, Mass., 1991.

c. What is the behavior of Y as X tends to infinity?
d. Can you give an example where such a model may be appropriate?

6.4. Consider the log-linear model:

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

Plot Y on the vertical axis and X on the horizontal axis. Draw the curves showing the relationship between Y and X when $\beta_2 = 1$, and when $\beta_2 > 1$, and when $\beta_2 < 1$.

6.5. Consider the following models:

$$\text{Model I:} \quad Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\text{Model II:} \quad Y_i^* = \alpha_1 + \alpha_2 X_i^* + u_i$$

where Y^* and X^* are standardized variables. Show that $\hat{\alpha}_2 = \hat{\beta}_2(S_x/S_y)$ and hence *establish that although the regression slope coefficients are independent of the change of origin they are not independent of the change of scale.*

6.6. Consider the following models:

$$\ln Y_i^* = \alpha_1 + \alpha_2 \ln X_i^* + u_i^*$$
$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

where $Y_i^* = w_1 Y_i$ and $X_i^* = w_2 X_i$, the w's being constants.

a. Establish the relationships between the two sets of regression coefficients and their standard errors.
b. Is the r^2 different between the two models?

6.7. Between regressions (6.6.8) and (6.6.10), which model do you prefer? Why?

6.8. For the regression (6.6.8), test the hypothesis that the slope coefficient is not significantly different from 0.005.

6.9. From the Phillips curve given in (6.7.3), is it possible to estimate the natural rate of unemployment? How?

6.10. The Engel expenditure curve relates a consumer's expenditure on a commodity to his or her total income. Letting Y = consumption expenditure on a commodity and X = consumer income, consider the following models:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$Y_i = \beta_1 + \beta_2 (1/X_i) + u_i$$
$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i$$
$$\ln Y_i = \ln \beta_1 + \beta_2 (1/X_i) + u_i$$
$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

Which of these model(s) would you choose for the Engel expenditure curve and why? (*Hint:* Interpret the various slope coefficients, find out the expressions for elasticity of expenditure with respect to income, etc.)

6.11. Consider the following model:

$$Y_i = \frac{e^{\beta_1+\beta_2 X_i}}{1 + e^{\beta_1+\beta_2 X_i}}$$

As it stands, is this a linear regression model? If not, what "trick," if any, can you use to make it a linear regression model? How would you interpret the resulting model? Under what circumstances might such a model be appropriate?

6.12. Graph the following models (for ease of exposition, we have omitted the observation subscript, i):

a. $Y = \beta_1 X^{\beta_2}$, for $\beta_2 > 1, \beta_2 = 1, 0 < \beta_2 < 1, \ldots$.

b. $Y = \beta_1 e^{\beta_2 X}$, for $\beta_2 > 0$ and $\beta_2 < 0$.

Discuss where such models might be appropriate.

Problems

6.13. You are given the data in Table 6.7.* Fit the following model to these data and obtain the usual regression statistics and interpret the results:

$$\frac{100}{100 - Y_i} = \beta_1 + \beta_2\left(\frac{1}{X_i}\right)$$

TABLE 6.7

Y_i	86	79	76	69	65	62	52	51	51	48
X_i	3	7	12	17	25	35	45	55	70	120

6.14. To measure the elasticity of substitution between capital and labor inputs Arrow, Chenery, Minhas, and Solow, the authors of the now famous CES (constant elasticity of substitution) production function, used the following model†:

$$\log\left(\frac{V}{L}\right) = \log \beta_1 + \beta_2 \log W + u$$

where (V/L) = value added per unit of labor
L = labor input
W = real wage rate

The coefficient β_2 measures the elasticity of substitution between labor and capital (i.e., proportionate change in factor proportions/proportionate change in relative factor prices). From the data given in Table 6.8, verify that the estimated elasticity is 1.3338 and that it is not statistically significantly different from 1.

6.15. Table 6.9 gives data on the GDP (gross domestic product) deflator for domestic goods and the GDP deflator for imports for Singapore for the period 1968–1982. The GDP deflator is often used as an indicator of inflation in place of the CPI. Singapore is a small, open economy, heavily dependent on foreign trade for its survival.

Source: Adapted from J. Johnston, *Econometric Methods,* 3d ed., McGraw-Hill, New York, 1984, p. 87. Actually this is taken from an econometric examination of Oxford University in 1975.

†"Capital-Labor Substitution and Economic Efficiency," *Review of Economics and Statistics,* August 1961, vol. 43, no. 5, pp. 225–254.

TABLE 6.8

Industry	log(*V*/*L*)	log *W*
Wheat flour	3.6973	2.9617
Sugar	3.4795	2.8532
Paints and varnishes	4.0004	3.1158
Cement	3.6609	3.0371
Glass and glassware	3.2321	2.8727
Ceramics	3.3418	2.9745
Plywood	3.4308	2.8287
Cotton textiles	3.3158	3.0888
Woolen textiles	3.5062	3.0086
Jute textiles	3.2352	2.9680
Chemicals	3.8823	3.0909
Aluminum	3.7309	3.0881
Iron and steel	3.7716	3.2256
Bicycles	3.6601	3.1025
Sewing machines	3.7554	3.1354

Source: Damodar Gujarati, "A Test of ACMS Production Function: Indian Industries, 1958," *Indian Journal of Industrial Relations,* vol. 2, no. 1, July 1966, pp. 95–97.

TABLE 6.9

Year	GDP deflator for domestic goods, *Y*	GDP deflator for imports, *X*
1968	1000	1000
1969	1023	1042
1970	1040	1092
1971	1087	1105
1972	1146	1110
1973	1285	1257
1974	1485	1749
1975	1521	1770
1976	1543	1889
1977	1567	1974
1978	1592	2015
1979	1714	2260
1980	1841	2621
1981	1959	2777
1982	2033	2735

Source: Colin Simkin, "Does Money Matter in Singapore?" *The Singapore Economic Review,* vol. XXIX, no.1, April 1984, Table 6, p. 8.

To study the relationship between domestic and world prices, you are given the following models:

$$\textbf{1.}\quad Y_t = \alpha_1 + \alpha_2 X_t + u_t$$

$$\textbf{2.}\quad Y_t = \beta_2 X_t + u_t$$

where Y = GDP deflator for domestic goods and X = GDP deflator for imports.

a. How would you choose between the two models a priori?
b. Fit both models to the data and decide which gives a better fit.
c. What other model(s) might be appropriate for the data?

6.16. Refer to the data given in exercise 6.15. The means of Y and X are 1456 and 1760, respectively, and the corresponding standard deviations are 346 and 641. Estimate the following regression:

$$Y_t^* = \alpha_1 + \alpha_2 X_t^* + u_t$$

where the starred variables are standardized variables, and interpret the results.

6.17. Refer to Table 6.3. Find out the rate of growth of expenditure on durable goods. What is the estimated *semielasticity?* Interpret your results. Would it make sense to run a double-log regression with expenditure on durable goods as the regressand and time as the regressor? How would you interpret the slope coefficient in this case.

6.18. From the data given in Table 6.3, find out the growth rate of expenditure on nondurable goods and compare your results with those obtained from problem 6.17.

6.19. Revisit exercise 1.7. Now that you know several functional forms, which one might be appropriate to study the relationship between advertising impressions retained and the amount of money spent on advertising? Show the necessary calculations.

APPENDIX 6A

6A.1 DERIVATION OF LEAST-SQUARES ESTIMATORS FOR REGRESSION THROUGH THE ORIGIN

We want to minimize

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_2 X_i)^2 \tag{1}$$

with respect to $\hat{\beta}_2$.

Differentiating (1) with respect to $\hat{\beta}_2$, we obtain

$$\frac{d \sum \hat{u}_i^2}{d\hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_2 X_i)(-X_i) \tag{2}$$

Setting (2) equal to zero and simplifying, we get

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad \textbf{(6.1.6) = (3)}$$

Now substituting the PRF: $Y_i = \beta_2 X_i + u_i$ into this equation, we obtain

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum X_i(\beta_2 X_i + u_i)}{\sum X_i^2} \\ &= \beta_2 + \frac{\sum X_i u_i}{\sum X_i^2} \end{aligned} \tag{4}$$

[*Note:* $E(\hat{\beta}_2) = \beta_2$.] Therefore,

$$E(\hat{\beta}_2 - \beta_2)^2 = E\left[\frac{\sum X_i u_i}{\sum X_i^2}\right]^2 \tag{5}$$

Expanding the right-hand side of (5) and noting that the X_i are nonstochastic and the u_i are homoscedastic and uncorrelated, we obtain

$$\text{var}\,(\hat{\beta}_2) = E(\hat{\beta}_2 - \beta_2)^2 = \frac{\sigma^2}{\sum X_i^2} \qquad \textbf{(6.1.7) = (6)}$$

Incidentally, note that from (2) we get, after equating it to zero

$$\sum \hat{u}_i X_i = 0 \tag{7}$$

From Appendix 3A, Section 3A.1 we see that when the intercept term is present in the model, we get in addition to (7) the condition $\sum \hat{u}_i = 0$. From the mathematics just given it should be clear why the regression through the origin model may not have the error sum, $\sum \hat{u}_i$, equal to zero.

Suppose we want to impose the condition that $\sum \hat{u}_i = 0$. In that case we have

$$\begin{aligned} \sum Y_i &= \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \\ &= \hat{\beta}_2 \sum X_i, \qquad \text{since } \sum \hat{u}_i = 0 \text{ by construction} \end{aligned} \tag{8}$$

This expression then gives

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum Y_i}{\sum X_i} \\ &= \frac{\bar{Y}}{\bar{X}} = \frac{\text{mean value of } Y}{\text{mean value of } X} \end{aligned} \tag{9}$$

But this estimator is not the same as (3) above or (6.1.6). And since the $\hat{\beta}_2$ of (3) is unbiased (why?), the $\hat{\beta}_2$ of (9) cannot be unbiased.

The upshot is that, in regression through the origin, we cannot have both $\sum \hat{u}_i X_i$ and $\sum \hat{u}_i$ equal to zero, as in the conventional model. The only condition that is satisfied is that $\sum \hat{u}_i X_i$ is zero.

Recall that

$$Y_i = \hat{Y}_i + \hat{u} \qquad \textbf{(2.6.3)}$$

Summing this equation on both sides and dividing by N, the sample size, we obtain

$$\bar{Y} = \bar{\hat{Y}} + \bar{\hat{u}} \tag{10}$$

Since for the zero intercept model $\sum \hat{u}_i$ and, therefore $\bar{\hat{u}}$, need not be zero, it then follows that

$$\bar{Y} \neq \bar{\hat{Y}} \tag{11}$$

that is, the mean of actual Y values need not be equal to the mean of the estimated Y values; the two mean values are identical for the intercept-present model, as can be seen from (3.1.10).

It was noted that, for the zero-intercept model, r^2 can be negative, whereas for the conventional model it can never be negative. This condition can be shown as follows.

Using (3.5.5a), we can write

$$r^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \tag{12}$$

Now for the conventional, or intercept-present, model, Eq. (3.3.6) shows that

$$\text{RSS} = \sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \leq \sum y_i^2 \tag{13}$$

unless $\hat{\beta}_2$ is zero (i.e., X has no influence on Y whatsoever). That is, for the conventional model, RSS $\leq$ TSS, or, r^2 can never be negative.

For the zero-intercept model it can be shown analogously that

$$\text{RSS} = \sum \hat{u}_i^2 = \sum Y_i^2 - \hat{\beta}_2^2 \sum X_i^2 \tag{14}$$

(*Note:* The sums of squares of Y and X are not mean-adjusted.) Now there is no guarantee that this RSS will always be less than $\sum y_i^2 = \sum Y_i^2 - N\bar{Y}^2$ (the TSS), which suggests that RSS can be greater than TSS, implying that r^2, as conventionally defined, can be negative. Incidentally, notice that in this case RSS will be greater than TSS if $\hat{\beta}_2^2 \sum X_i^2 < N\bar{Y}^2$.

6A.2 PROOF THAT A STANDARDIZED VARIABLE HAS ZERO MEAN AND UNIT VARIANCE

Consider the random variable (r.v.) Y with the (sample) mean value of $\bar{Y}$ and (sample) standard deviation of S_y. Define

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_y} \tag{15}$$

Hence Y_i^* is a standardized variable. Notice that standardization involves a dual operation: (1) change of the origin, which is the numerator of (15), and (2) change of scale, which is the denominator. Thus, standardization involves both a change of the origin and change of scale.

Now

$$\bar{Y}_i^* = \frac{1}{S_y} \frac{\sum (Y_i - \bar{Y})}{n} = 0 \tag{16}$$

since the sum of deviation of a variable from its mean value is always zero. Hence the mean value of the standardized value is zero. (*Note:* We could pull out the S_y term from the summation sign because its value is known.)

Now

$$\begin{aligned} S_{y^*}^2 &= \sum \frac{(Y_i - \bar{Y})^2/(n-1)}{S_y^2} \\ &= \frac{1}{(n-1)S_y^2} \sum (Y_i - \bar{Y})^2 \\ &= \frac{(n-1)S_y^2}{(n-1)S_y^2} = 1 \end{aligned} \tag{17}$$

Note that

$$S_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1}$$

which is the sample variance of Y.