Until now we have considered models that were linear in parameters as well as in variables. But recall that in this textbook our concern is with models that are linear in parameters; the Y and X variables do not necessarily have to be linear. As a matter of fact, as we show in this chapter, there are many economic phenomena for which the linear-in-parameters/linear-in-variables (LIP/LIV, for short) regression models may not be adequate or appropriate.

For example, suppose for the LIP/LIV math S.A.T. score function given in Equation (2.20) we want to estimate the score elasticity of the math S.A.T., that is, the percentage change in the math S.A.T. score for a percentage change in annual family income. We cannot estimate this elasticity from Eq. (2.20) directly because the slope coefficient of that model simply gives the absolute change in the (average) math S.A.T. score for a unit (say, a dollar) change in the annual family income, but this is not elasticity. Such elasticity, however, can be readily computed from the so-called log-linear models that will be discussed in Section 5.1. As we will show, this model, although linear in the parameters, is not linear in the variables.

For another example, suppose we want to find out the rate of growth\(^1\) over time of an economic variable, such as gross domestic product (GDP) or money supply, or unemployment rate. As we show in Section 5.4, this growth rate can

---

\(^1\)If \(Y_t\) and \(Y_{t-1}\) are values of a variable, say, GDP, at time \(t\) and \((t - 1)\), say, 2009 and 2008, then the rate of growth of \(Y\) in the two time periods is measured as \(\frac{Y_t - Y_{t-1}}{Y_t} \cdot 100\), which is simply the relative, or proportional, change in \(Y\) multiplied by 100. It is shown in Section 5.4 how the semilog model can be used to measure the growth rate over a longer period of time.
be measured by the so-called *semilog model* which, while linear in parameters, is nonlinear in variables.

Note that even within the confines of the linear-in-parameter regression models, a regression model can assume a variety of *functional forms*. In particular, in this chapter we will discuss the following types of regression models:

1. **Log-linear or constant elasticity models** (Section 5.1).
2. **Semilog models** (Sections 5.4 and 5.5).
3. **Reciprocal models** (Section 5.6).
4. **Polynomial regression models** (Section 5.7).
5. **Regression-through-the-origin, or zero intercept, model** (Section 5.8).

An important feature of all these models is that they are linear in parameters (or can be made so by simple algebraic manipulations), but they are not necessarily linear in variables. In Chapter 2 we discussed the technical meaning of linearity in both variables and parameters. Briefly, for a regression model linear in explanatory variable(s) the rate of change (i.e., the slope) of the dependent variable remains constant for a unit change in the explanatory variable, whereas for regression models nonlinear in explanatory variable(s) the slope does not remain constant.

To introduce the basic concepts, and to illustrate them graphically, initially we will consider two-variable models and then extend the discussion to multiple regression models.

### 5.1 HOW TO MEASURE ELASTICITY: THE LOG-LINEAR MODEL

Let us revisit our math S.A.T. score function discussed in Chapters 2 and 3. But now consider the following model for the math S.A.T. score function. (To ease the algebra, we will introduce the error term \( u_i \) later.)

\[
Y_i = AX_i^{B_2}
\]  

(5.1)

where \( Y \) is math S.A.T. score and \( X \) is annual family income.

This model is nonlinear in the variable \( X \).\(^2\) Let us, however, express Equation (5.1) in an alternative, but equivalent, form, as follows:

\[
\ln Y_i = \ln A + B_2 \ln X_i
\]  

(5.2)

\(^2\)Using calculus, it can be shown that

\[
\frac{dY}{dX} = AB_2X^{(B_2-1)}
\]

which shows that the rate of change of \( Y \) with respect to \( X \) is not independent of \( X \); that is, it is not constant. By definition, then, model (5.1) is not linear in variable \( X \).
where ln = the natural log, that is, logarithm to the base \( e \).\(^3\) Now if we let

\[ B_1 = \ln A \]  
\[ (5.3) \]

we can write Equation (5.2) as

\[ \ln Y_i = B_1 + B_2 \ln X_i \]
\[ (5.4) \]

And for estimating purposes, we can write this model as

\[ \ln Y_i = B_1 + B_2 \ln X_i + u_i \]
\[ (5.5) \]

This is a linear regression model, for the parameters \( B_1 \) and \( B_2 \) enter the model linearly.\(^4\) It is of interest that this model is also linear in the logarithms of the variables \( Y \) and \( X \). (Note: The original model [5.1] was nonlinear in \( X \).) Because of this linearity, models like Equation (5.5) are called double-log (because both variables are in the log form) or log-linear (because of linearity in the logs of the variables) models.

Notice how an apparently nonlinear model (5.1) can be converted into a linear (in the parameter) model by suitable transformation, here the logarithmic transformation. Now letting \( Y_i^* = \ln Y_i \) and \( X_i^* = \ln X_i \), we can write model (5.5) as

\[ Y_i^* = B_1 + B_2 X_i^* + u_i \]
\[ (5.6) \]

which resembles the models we have considered in previous chapters; it is linear in both the parameters and the transformed variables \( Y^* \) and \( X^* \).

If the assumptions of the classical linear regression model (CLRM) are satisfied for the transformed model, regression (5.6) can be estimated easily with the usual ordinary least squares (OLS) routine and the estimators thus obtained will have the usual best linear unbiased estimator (BLUE) property.\(^5\)

One attractive feature of the double-log, or log-linear, model that has made it popular in empirical work is that the slope coefficient \( B_2 \) measures the elasticity of \( Y \) with respect to \( X \), that is, the percentage change in \( Y \) for a given (small) percentage change in \( X \).

\(^3\)Appendix 5A discusses logarithms and their properties for the benefit of those who need it.

\(^4\)Note that since \( B_1 = \ln A \), \( A \) can be expressed as \( A = \text{antilog} (B_1) \) which is, mathematically speaking, a nonlinear transformation. In practice, however, the intercept \( A \) often does not have much concrete meaning.

\(^5\)Any regression package now routinely computes the logs of (positive) numbers. So there is no additional computational burden involved.
Symbolically, if we let $\Delta Y$ stand for a small change in $Y$ and $\Delta X$ for a small change in $X$, we define the elasticity coefficient, $E$, as

$$E = \frac{\% \text{ change in } Y}{\% \text{ change in } X} = \frac{\Delta Y / Y \cdot 100}{\Delta X / X \cdot 100} = \frac{\Delta Y / X}{\Delta X / Y} = \text{slope} \left( \frac{X}{Y} \right).$$

Thus, if $Y$ represents the quantity of a commodity demanded and $X$ its unit price, $B_2$ measures the price elasticity of demand.

All this can be shown graphically.

Figure 5-1(a) represents the function (5.1), and Figure 5-1(b) shows its logarithmic transformation. The slope of the straight line shown in Figure 5-1(b) gives the estimate of price elasticity, $-B_2$. An important feature of the log-linear model should be apparent from Figure 5-1(b). Since the regression line is a straight line (in the logs of $Y$ and $X$), its slope $(-B_2)$ is constant throughout. And since this slope coefficient is equal to the elasticity coefficient, for this

$$E = \frac{dY}{dX} \cdot \frac{X}{Y}.$$

In calculus notation

$$E = \frac{dY}{dX} \cdot \frac{X}{Y},$$

where $dY/dX$ means the derivative of $Y$ with respect to $X$, that is, the rate of change of $Y$ with respect to $X$. $\Delta Y/\Delta X$ is an approximation of $dY/dX$. Note: For the transformed model (5.6),

$$B_2 = \frac{\Delta Y}{\Delta X} = \frac{\Delta \ln Y}{\Delta \ln X} = \frac{\Delta Y / Y}{\Delta X / X} = \frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y},$$

which is the elasticity of $Y$ with respect to $X$ as per Equation (5.7). As noted in Appendix 5A, a change in the log of a number is a relative or proportional change. For example, $\Delta \ln Y = \frac{\Delta Y}{Y}$.
Because of this special feature, the double-log or log-linear model is also known as the \textit{constant elasticity model}. Therefore, we will use all of these terms interchangeably.

\textbf{Example 5.1 Math S.A.T. Score Function Revisited}

In Equation (3.46) we presented the linear (in variables) function for our math S.A.T. score example. Recall, however, that the scattergram showed that the relationship between math S.A.T. scores and annual family income was \textit{approximately} linear because not all points were really on a straight line. Eq. (3.46) was, of course, developed for pedagogy. Let us see if the log-linear model fits the data given in Table 2-2, which for convenience is reproduced in Table 5-1.

The OLS regression based on the log-linear data gave the following results:

\begin{equation}
\ln Y_i = 4.8877 + 0.1258 \ln X_i \\
se = (0.1573) \quad (0.0148) \\
t = (31.0740) \quad (8.5095) \\
p = (1.25 \times 10^{-9})(2.79 \times 10^{-5}) \quad r^2 = 0.9005
\end{equation}

As these results show, the (constant) score elasticity is \( \approx 0.13 \), suggesting that if the annual family income increases by 1 percent, the math S.A.T. score on average increases \( \approx 0.13 \) percent. By convention, an elasticity coefficient less than 1 indicates that the relationship is \textit{elastic}.

\textsuperscript{7}Note carefully, however, that in general, elasticity and slope coefficients are different concepts. As Eq. (5.7) makes clear, elasticity is equal to the slope times the ratio of \( X/Y \). It is only for the double-log, or log-linear, model that the two are identical.

\textit{The Pink Professor}
than 1 in absolute value is said to be inelastic, whereas if it is greater than 1, it is called elastic. An elasticity coefficient of 1 (in absolute value) has unitary elasticity. Therefore, in our example, the math S.A.T. score is inelastic; the math score increases proportionately less than the increase in annual family income.

The interpretation of the intercept of \( \approx 4.89 \) means that the average value of \( \ln Y \) is 4.89 if the value of \( \ln X \) is zero. Again, this mechanical interpretation of the intercept may not have concrete economic meaning.8

The interpretation of \( r^2 = 0.9005 \) is that \( \approx 90 \) percent of the variation in the log of \( Y \) is explained by the variation in the log of \( X \).

The regression line in Equation (5.8) is sketched in Figure 5-2. Notice that this figure is quite similar to Figure 2-1.

**Hypothesis Testing in Log-Linear Models**

There is absolutely no difference between the linear and log-linear models insofar as hypothesis testing is concerned. Under the assumption that the error term follows the normal distribution with mean zero and constant variance \( \sigma^2 \), it follows that each estimated regression coefficient is normally distributed. Or, if we replace \( \sigma^2 \) by its unbiased estimator \( \hat{\sigma}^2 \), each estimator follows the \( t \) distribution with degrees of freedom (d.f.) equal to \((n - k)\), where \( k \) is the number of parameters.

8Since \( \ln Y = 4.8877 \) when \( \ln X \) is zero, if we take the antilog of this number, we obtain \( \approx 132.94 \). Thus, the average math S.A.T. score is about 133 points if the log of annual family income is zero. For the linear model given in Eq. (3.46), the intercept value was about 432.41 points when annual family income (not the log of income) was zero.

*The Pink Professor*
estimated, including the intercept. In the two-variable case, \( k = 2 \), in the three-variable case, \( k = 3 \), etc.

From the regression (5.8), you can readily check that the slope coefficient is statistically significantly different from zero since the \( t \) value of \( 8.51 \) has a \( p \) value of \( 2.79 \times 10^{-3} \), which is very small. If the null hypothesis that annual family income has no relationship to math S.A.T. score were true, our chances of obtaining a \( t \) value of as much as 8.51 or greater would be about 3 in 100,000! The intercept value of 4.8877 is also statistically significant because the \( p \) value is about \( 1.25 \times 10^{-9} \).

### 5.2 COMPARING LINEAR AND LOG-LINEAR REGRESSION MODELS

We take this opportunity to consider an important practical question. We have fitted a linear (in variables) S.A.T. score function, Eq. (3.46), as well as a log-linear function, Eq. (5.8), for our S.A.T. score example. Which model should we choose? Although it may seem logical that students with higher family income would tend to have higher S.A.T. scores, indicating a positive relationship, we don't really know which particular functional form defines the relationship between them. That is, we may not know if we should fit the linear, log-linear, or some other model. The functional form of the regression model then becomes essentially an empirical question. Are there any guidelines or rules of thumb that we can follow in choosing among competing models?

One guiding principle is to plot the data. If the scattergram shows that the relationship between two variables looks reasonably linear (i.e., a straight line), the linear specification might be appropriate. But if the scattergram shows a nonlinear relationship, plot the log of \( Y \) against the log of \( X \). If this plot shows an approximately linear relationship, a log-linear model may be appropriate. Unfortunately, this guiding principle works only in the simple case of two-variable regression models and is not very helpful once we consider multiple regressions; it is not easy to draw scattergrams in multiple dimensions. We need other guidelines.

Why not choose the model on the basis of \( r^2 \); that is, choose the model that gives the highest \( r^2 \)? Although intuitively appealing, this criterion has its own problems. First, as noted in Chapter 4, to compare the \( r^2 \) values of two models, the dependent variable must be in the same form. For model (3.46), the dependent variable is \( Y \), whereas for the model (5.8), it is \( \ln(Y) \), and these two dependent variables are obviously not the same. Therefore, \( r^2 = 0.7869 \) of the linear model (3.46) and \( r^2 = 0.9005 \) of the log-linear model are not directly comparable, even though they are approximately the same in the present case.

---

9 A cautionary note here: Remember that regression models do not imply causation, so we are not implying that having a higher annual family income causes higher math S.A.T. scores, only that we would tend to see the two together. There may be several other reasons explaining this result. Perhaps students with higher family incomes are able to afford S.A.T. preparation classes or attend schools that focus more on material typically covered in the exam.

10 It does not matter what form the independent or explanatory variables take; they may or may not be linear.

---

The Pink Professor
The reason that we cannot compare these two $r^2$ values is not difficult to grasp. By definition, $r^2$ measures the proportion of the variation in the dependent variable explained by the explanatory variable(s). In the linear model (3.46) $r^2$ thus measures the proportion of the variation in $Y$ explained by $X$, whereas in the log-linear model (5.8) it measures the proportion of the variation in the log of $Y$ explained by the log of $X$. Now the variation in $Y$ and the variation in the log of $Y$ are conceptually different. The variation in the log of a number measures the relative or proportional change (or percentage change if multiplied by 100), and the variation in a number measures the absolute change. Thus, for the linear model (3.46), $\approx 79$ percent of the variation in $Y$ is explained by $X$, whereas for the log-linear model, $\approx 90$ percent of the variation in the log of $Y$ is explained by the log of $X$. If we want to compare the two $r^2$s, we can use the method discussed in Problem 5.16.

Even if the dependent variable in the two models is the same so that two $r^2$ values can be directly compared, you are well-advised against choosing a model on the basis of a high $r^2$ value criterion. This is because, as pointed out in Chapter 4, an $r^2 (=R^2)$ can always be increased by adding more explanatory variables to the model. Rather than emphasizing the $r^2$ value of a model, you should consider factors such as the relevance of the explanatory variables included in the model (i.e., the underlying theory), the expected signs of the coefficients of the explanatory variables, their statistical significance, and certain derived measures like the elasticity coefficient. These should be the guiding principles in choosing between two competing models. If based on these criteria one model is preferable to the other, and if the chosen model also happens to have a higher $r^2$ value, then well and good. But avoid the temptation of choosing a model only on the basis of the $r^2$ value alone.

Comparing the results of the log-linear score function (5.8) versus the linear function (3.46), we observe that in both models the slope coefficient is positive, as per prior expectations. Also, both slope coefficients are statistically significant. However, we cannot compare the two slope coefficients directly, for in the LIV model it measures the absolute rate of change in the dependent variable, whereas in the log-linear model it measures elasticity of $Y$ with respect to $X$.

If for the LIV model we can measure score elasticity, then it is possible to compare the two slope coefficients. To do this, we can use Equation (5.7), which shows that elasticity is equal to the slope times the ratio of $X$ to $Y$. Although for the linear model the slope coefficient remains the same (Why?), which is 0.0013 in our S.A.T. score example, the elasticity changes from point to point on the linear curve because the ratio $X/Y$ changes from point to point. From Table 5-1 we see that there are 10 different math S.A.T. score and annual family income figures. Therefore, in principle we can compute 10 different elasticity coefficients. In practice, however, the elasticity coefficient for the

---

11If a number goes from 45 to 50, the absolute change is 5, but the relative change is $(50 - 45)/45 \approx 0.1111$, or about 11.11 percent.

The Pink Professor
A linear model is often computed at the sample mean values of $X$ and $Y$ to obtain a measure of average elasticity. That is,

\[
\text{Average elasticity} = \frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y}
\]  

(5.9)

where $\bar{X}$ and $\bar{Y}$ are sample mean values. For the data given in Table 5-1, $\bar{X} = 56,000$ and $\bar{Y} = 507$. Thus, the average elasticity for our sample is

\[
\text{Average score elasticity} = (0.0013) \frac{56,000}{507} = 0.1436
\]

It is interesting to note that for the log-linear function the score elasticity coefficient was 0.1258, which remains the same no matter at what income the elasticity is measured (see Figure 5-1[b]). This is why such a model is called a constant elasticity model. For the LIV, on the other hand, the elasticity coefficient changes from point to point on the score = family income curve.\(^\text{12}\)

The fact that for the linear model the elasticity coefficient changes from point to point and that for the log-linear model it remains the same at all points on the demand curve means that we have to exercise some judgment in choosing between the two specifications, for, in practice, both these assumptions may be extreme. It is possible that over a small segment of the expenditure curve the elasticity remains constant but that over some other segment(s) it is variable.

5.3 MULTIPLE LOG-LINEAR REGRESSION MODELS

The two-variable log-linear model can be generalized easily to models containing more than one explanatory variable. For example, a three-variable log-linear model can be expressed as

\[
\ln Y_i = B_1 + B_2 \ln X_{2i} + B_3 \ln X_{3i} + u_i
\]

(5.10)

In this model the partial slope coefficients $B_2$ and $B_3$ are also called the partial elasticity coefficients.\(^\text{13}\) Thus, $B_2$ measures the elasticity of $Y$ with respect to $X_2$, holding the influence of $X_3$ constant; that is, it measures the percentage change in $Y$ for a percentage change in $X_2$ holding the influence of $X_3$ constant. Since the influence of $X_3$ is held constant, it is called a partial elasticity. Similarly, $B_3$

\(^\text{12}\)Notice this interesting fact: For the LIV model, the slope coefficient is constant but the elasticity coefficient is variable. However, for the log-linear model, the elasticity coefficient is constant but the slope coefficient is variable, which can be seen at once from the formula given in footnote 2.

\(^\text{13}\)The calculus-minded reader will recognize that the partial derivative of $\ln Y$ with respect to $\ln X_2$ is

\[
B_2 = \frac{\partial \ln Y}{\partial \ln X_2} = \frac{\partial Y / Y}{\partial X_2 / X_2} = \frac{\partial Y}{\partial X_2} \cdot \frac{X_2}{Y}
\]

which by definition is elasticity of $Y$ with respect to $X_2$. Likewise, $B_3$ is the elasticity of $Y$ with respect to $X_3$. 

The Pink Professor
measures the (partial) elasticity of $Y$ with respect to $X_3$, holding the influence of $X_2$ constant. In short, in a multiple log-linear model, each partial slope coefficient measures the partial elasticity of the dependent variable with respect to the explanatory variable in question, holding all other variables constant.

**Example 5.2. The Cobb-Douglas Production Function**

As an example of model (5.10), let $Y =$ output, $X_2 =$ labor input, and $X_3 =$ capital input. In that case model (5.10) becomes a production function—a function that relates output to labor and capital inputs. As a matter of fact, regression (5.10) in this case represents the celebrated **Cobb-Douglas (C-D) production function**. As an illustration, consider the data given in Table 5-2, which relates to Mexico for the years 1955 to 1974. $Y$, the output, is measured by gross domestic product (GDP) (millions of 1960 pesos), $X_2$, the labor input, is measured by total employment (thousands of people), and $X_3$, the capital input, is measured by stock of fixed capital (millions of 1960 pesos).

**TABLE 5-2 REAL GDP, EMPLOYMENT, AND REAL FIXED CAPITAL, MEXICO, 1955–1974**

<table>
<thead>
<tr>
<th>Year</th>
<th>GDPa</th>
<th>Employmentb</th>
<th>Fixed capitalc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955</td>
<td>114043</td>
<td>8310</td>
<td>182113</td>
</tr>
<tr>
<td>1956</td>
<td>120410</td>
<td>8529</td>
<td>193745</td>
</tr>
<tr>
<td>1957</td>
<td>129187</td>
<td>8738</td>
<td>205192</td>
</tr>
<tr>
<td>1958</td>
<td>134705</td>
<td>8952</td>
<td>215130</td>
</tr>
<tr>
<td>1959</td>
<td>139960</td>
<td>9171</td>
<td>225021</td>
</tr>
<tr>
<td>1960</td>
<td>150511</td>
<td>9569</td>
<td>237026</td>
</tr>
<tr>
<td>1961</td>
<td>157897</td>
<td>9527</td>
<td>248897</td>
</tr>
<tr>
<td>1962</td>
<td>165286</td>
<td>9662</td>
<td>260661</td>
</tr>
<tr>
<td>1963</td>
<td>178491</td>
<td>10334</td>
<td>275466</td>
</tr>
<tr>
<td>1964</td>
<td>199457</td>
<td>10981</td>
<td>295378</td>
</tr>
<tr>
<td>1965</td>
<td>212323</td>
<td>11746</td>
<td>315715</td>
</tr>
<tr>
<td>1966</td>
<td>226977</td>
<td>11521</td>
<td>337642</td>
</tr>
<tr>
<td>1967</td>
<td>241194</td>
<td>11540</td>
<td>363599</td>
</tr>
<tr>
<td>1968</td>
<td>260881</td>
<td>12066</td>
<td>391847</td>
</tr>
<tr>
<td>1969</td>
<td>277498</td>
<td>12297</td>
<td>422382</td>
</tr>
<tr>
<td>1970</td>
<td>296530</td>
<td>12955</td>
<td>455049</td>
</tr>
<tr>
<td>1971</td>
<td>306712</td>
<td>13338</td>
<td>484677</td>
</tr>
<tr>
<td>1972</td>
<td>329030</td>
<td>13738</td>
<td>520553</td>
</tr>
<tr>
<td>1973</td>
<td>354057</td>
<td>15924</td>
<td>561531</td>
</tr>
<tr>
<td>1974</td>
<td>374977</td>
<td>14154</td>
<td>609825</td>
</tr>
</tbody>
</table>

**Notes:**
* $a$Millions of 1960 pesos.
* $b$Thousands of people.
* $c$Millions of 1960 pesos.

Based on the data given in Table 5-2, the following results were obtained using the MINITAB statistical package:

\[
\ln Y_t = -1.6524 + 0.3397 \ln X_{2t} + 0.8460 \ln X_{3t}
\]

<table>
<thead>
<tr>
<th>se</th>
<th>t</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.6062)</td>
<td>(-2.73)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>(0.1857)</td>
<td>(1.83)</td>
<td>(0.085)</td>
</tr>
<tr>
<td>(0.09343)</td>
<td>(9.06)</td>
<td>(0.000)*</td>
</tr>
</tbody>
</table>

\[R^2 = 0.995\]
\[F = 1719.23\] (0.000)**

The interpretation of regression (5.11) is as follows. The partial slope coefficient of 0.3397 measures the elasticity of output with respect to the labor input. Specifically, this number states that, holding the capital input constant, if the labor input increases by 1 percent, on the average, output goes up by about 0.34 percent. Similarly, holding the labor input constant, if the capital input increases by 1 percent, on the average, output goes up by about 0.85 percent. If we add the elasticity coefficients, we obtain an economically important parameter, called the returns to scale parameter, which gives the response of output to a proportional change in inputs. If the sum of the two elasticity coefficients is 1, we have constant returns to scale (i.e., doubling the inputs simultaneously doubles the output); if it is greater than 1, we have increasing returns to scale (i.e., doubling the inputs simultaneously more than doubles the output); if it is less than 1, we have decreasing returns to scale (i.e., doubling the inputs less than doubles the output).

For Mexico, for the study period, the sum of the two elasticity coefficients is 1.1857, suggesting that perhaps the Mexican economy was characterized by increasing returns to scale.

Returning to the estimated coefficients, we see that both labor and capital are individually statistically significant on the basis of the one-tail test although the impact of capital seems to be more important than that of labor. (Note: We use a one-tail test because both labor and capital are expected to have a positive effect on output.)

The estimated F value is so highly significant (because the p value is almost zero) we can strongly reject the null hypothesis that labor and capital together do not have any impact on output.

The \(R^2\) value of 0.995 means that about 99.5 percent of the variation in the (log) of output is explained by the (logs) of labor and capital, a very high degree of explanation, suggesting that the model (5.11) fits the data very well.

**Example 5.3. The Demand for Energy**

Table 5-3 gives data on the indexes of aggregate final energy demand \(Y\), real GDP \(X_2\), and real energy price \(X_3\) for seven OECD countries (the

*Denotes extremely small value.
** p value of F, also extremely small.
TABLE 5-3 ENERGY DEMAND IN OECD COUNTRIES, 1960–1982

<table>
<thead>
<tr>
<th>Year</th>
<th>Final demand</th>
<th>Real GDP</th>
<th>Real energy price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>54.1</td>
<td>54.1</td>
<td>111.9</td>
</tr>
<tr>
<td>1961</td>
<td>55.4</td>
<td>56.4</td>
<td>112.4</td>
</tr>
<tr>
<td>1962</td>
<td>58.5</td>
<td>59.4</td>
<td>111.1</td>
</tr>
<tr>
<td>1963</td>
<td>61.7</td>
<td>62.1</td>
<td>110.2</td>
</tr>
<tr>
<td>1964</td>
<td>63.6</td>
<td>65.9</td>
<td>109.0</td>
</tr>
<tr>
<td>1965</td>
<td>66.8</td>
<td>69.5</td>
<td>108.3</td>
</tr>
<tr>
<td>1966</td>
<td>70.3</td>
<td>73.2</td>
<td>105.3</td>
</tr>
<tr>
<td>1967</td>
<td>73.5</td>
<td>75.7</td>
<td>105.4</td>
</tr>
<tr>
<td>1968</td>
<td>78.3</td>
<td>79.9</td>
<td>104.3</td>
</tr>
<tr>
<td>1969</td>
<td>83.8</td>
<td>83.8</td>
<td>101.7</td>
</tr>
<tr>
<td>1970</td>
<td>88.9</td>
<td>86.2</td>
<td>97.7</td>
</tr>
<tr>
<td>1971</td>
<td>91.8</td>
<td>89.8</td>
<td>100.3</td>
</tr>
<tr>
<td>1972</td>
<td>97.2</td>
<td>94.3</td>
<td>98.6</td>
</tr>
<tr>
<td>1973</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>1974</td>
<td>97.4</td>
<td>101.4</td>
<td>120.1</td>
</tr>
<tr>
<td>1975</td>
<td>93.5</td>
<td>100.5</td>
<td>131.0</td>
</tr>
<tr>
<td>1976</td>
<td>99.1</td>
<td>105.3</td>
<td>129.6</td>
</tr>
<tr>
<td>1977</td>
<td>100.9</td>
<td>109.9</td>
<td>137.7</td>
</tr>
<tr>
<td>1978</td>
<td>103.9</td>
<td>114.4</td>
<td>133.7</td>
</tr>
<tr>
<td>1979</td>
<td>106.9</td>
<td>118.3</td>
<td>144.5</td>
</tr>
<tr>
<td>1980</td>
<td>101.2</td>
<td>119.6</td>
<td>179.0</td>
</tr>
<tr>
<td>1981</td>
<td>98.1</td>
<td>121.1</td>
<td>189.4</td>
</tr>
<tr>
<td>1982</td>
<td>95.6</td>
<td>120.6</td>
<td>190.9</td>
</tr>
</tbody>
</table>

*Denotes extremely small value.

United States, Canada, Germany, France, the United Kingdom, Italy, and Japan) for the period 1960 to 1982. All indexes are with base $1973 = 100$. Using the data given in Table 5-3 and MINITAB we obtained the following log-linear energy demand function:

\[
\ln Y_t = 1.5495 + 0.9972 \ln X_{2t} - 0.3315 \ln X_{3t}
\]

\[
\begin{align*}
\text{se} & = (0.0903) \quad (0.0191) \quad (0.0243) \\
\text{t} & = (17.17) \quad (52.09) \quad (13.61) \\
\text{p value} & = (0.000)* \quad (0.000)* \quad (0.000)* \\
R^2 & = 0.994 \\
\overline{R^2} & = 0.994 \\
F & = 1688
\end{align*}
\]

As this regression shows, energy demand is positively related to income (as measured by real GDP) and negatively related to real price; these findings

\[\text{(5.12)}\]
accord with economic theory. The estimated income elasticity is about 0.99, meaning that if real income goes up by 1 percent, the average amount of energy demanded goes up by about 0.99 percent, or just about 1 percent, ceteris paribus. Likewise, the estimated price elasticity is about $-0.33$, meaning that, holding other factors constant, if energy price goes up by 1 percent, the average amount of energy demanded goes down by about 0.33 percent. Since this coefficient is less than 1 in absolute value, we can say that the demand for energy is price inelastic, which is not very surprising because energy is a very essential item for consumption.

The $R^2$ values, both adjusted and unadjusted, are very high. The $F$ value of about 1688 is also very high; the probability of obtaining such an $F$ value, if in fact $B_2 = B_3 = 0$ is true, is almost zero. Therefore, we can say that income and energy price together strongly affect energy demand.

### 5.4 HOW TO MEASURE THE GROWTH RATE: THE SEMILOG MODEL

As noted in the introduction to this chapter, economists, businesspeople, and the government are often interested in finding out the rate of growth of certain economic variables. For example, the projection of the government budget deficit (surplus) is based on the projected rate of growth of the GDP, the single most important indicator of economic activity. Likewise, the Fed keeps a strong eye on the rate of growth of consumer credit outstanding (auto loans, installment loans, etc.) to monitor its monetary policy.

In this section we will show how regression analysis can be used to measure such growth rates.

**Example 5.4. The Growth of the U.S. Population, 1975–2007**

Table 5-4 gives data on the U.S. population (in millions) for the period 1975 to 2007.

We want to measure the rate of growth of the U.S. population ($Y$) over this period. Now consider the following well-known compound interest formula from your introductory courses in money, banking, and finance:

$$Y_t = Y_0(1 + r)^t$$  \hspace{1cm} (5.13)$^{14}$

$Y_0 = $ the beginning, or initial, value of $Y$

$Y_t = $ $Y$‘s value at time $t$

$r = $ the compound (i.e., over time) rate of growth of $Y$

---

$^{14}$Suppose you deposit $Y_0 = $100 in a passbook account in a bank, paying, say, 6 percent interest per year. Here $r = 0.06$, or 6 percent. At the end of the first year this amount will grow to $Y_1 = 100(1 + 0.06) = 106$; at the end of the second year it will be $Y_2 = 106(1 + 0.06) = 112.36$ because in the second year you get interest not only on the initial $100 but also on the interest earned in the first year. In the third year this amount grows to $100(1 + 0.06)^3 = 119.1016$, etc.
Let us manipulate Equation (5.13) as follows. Take the (natural) log of Eq. (5.13) on both sides to obtain

\[ \ln Y = \ln Y_0 + t \ln(1 + r) \]  \hspace{1cm} (5.14)

Now let

\[ \quad B_1 = \ln Y_0 \] \hspace{1cm} (5.15)

\[ \quad B_2 = \ln(1 + r) \] \hspace{1cm} (5.16)

Therefore, we can express model (5.14) as

\[ \ln Y_t = B_1 + B_2 t \] \hspace{1cm} (5.17)

Now if we add the error term \( u_t \) to model (5.17), we will obtain\(^{15}\)

\[ \ln Y_t = B_1 + B_2 t + u_t \] \hspace{1cm} (5.18)

This model is like any other linear regression model in that parameters \( B_1 \) and \( B_2 \) are linear. The only difference is that the dependent variable is the logarithm of \( Y \) and the independent, or explanatory, variable is “time,” which will take values of 1, 2, 3, etc.

\(^{15}\)The reason we add the error term is that the compound interest formula will not exactly fit the data of Table 5-4.

\[ \text{Table 5-4} \quad \text{POPULATION OF UNITED STATES (MILLIONS OF PEOPLE),} \quad \text{1975–2007} \]

<table>
<thead>
<tr>
<th>U.S. population</th>
<th>Time</th>
<th>U.S. population</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>215.973</td>
<td>1</td>
<td>256.894</td>
<td>18</td>
</tr>
<tr>
<td>218.035</td>
<td>2</td>
<td>260.255</td>
<td>19</td>
</tr>
<tr>
<td>220.239</td>
<td>3</td>
<td>263.436</td>
<td>20</td>
</tr>
<tr>
<td>222.585</td>
<td>4</td>
<td>266.557</td>
<td>21</td>
</tr>
<tr>
<td>225.055</td>
<td>5</td>
<td>269.667</td>
<td>22</td>
</tr>
<tr>
<td>227.726</td>
<td>6</td>
<td>272.912</td>
<td>23</td>
</tr>
<tr>
<td>229.966</td>
<td>7</td>
<td>276.115</td>
<td>24</td>
</tr>
<tr>
<td>232.188</td>
<td>8</td>
<td>279.295</td>
<td>25</td>
</tr>
<tr>
<td>234.307</td>
<td>9</td>
<td>282.430</td>
<td>26</td>
</tr>
<tr>
<td>236.348</td>
<td>10</td>
<td>285.454</td>
<td>27</td>
</tr>
<tr>
<td>238.466</td>
<td>11</td>
<td>288.427</td>
<td>28</td>
</tr>
<tr>
<td>240.651</td>
<td>12</td>
<td>291.289</td>
<td>29</td>
</tr>
<tr>
<td>242.804</td>
<td>13</td>
<td>294.056</td>
<td>30</td>
</tr>
<tr>
<td>245.021</td>
<td>14</td>
<td>296.940</td>
<td>31</td>
</tr>
<tr>
<td>247.342</td>
<td>15</td>
<td>299.801</td>
<td>32</td>
</tr>
<tr>
<td>250.132</td>
<td>16</td>
<td>302.045</td>
<td>33</td>
</tr>
<tr>
<td>253.493</td>
<td>17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Note: 1975} = 1; \text{2007} = 33. \]

\[ \text{Source: Economic Report of the President, 2008, Table B34.} \]
Models like regression (5.18) are called **semilog models** because only one variable (in this case the dependent variable) appears in the logarithmic form. How do we interpret semilog models like regression (5.18)? Before we discuss this, note that model (5.18) can be estimated by the usual OLS method, assuming of course that the usual assumptions of OLS are satisfied. For the data of Table 5-4, we obtain the following regression results:

\[
\ln (\text{USpop}) = 5.3593 + 0.0107t \\
\]

The estimated regression line is sketched in Figure 5-3.

Note that in Eq. (5.19) we have only reported the \( t \) values.

The interpretation of regression (5.19) is as follows. The slope coefficient of 0.0107 means on the average the log of \( Y \) (U.S. population) has been increasing at the rate of 0.0107 per year. In plain English, \( Y \) has been increasing at the rate of 1.07 percent per year, for in a semilog model like regression (5.19) the slope coefficient measures the proportional or relative change in \( Y \) for a given absolute change in the explanatory variable, time in the present case. If this relative change is multiplied by 100, we obtain the percentage change or the growth.

Using calculus it can be shown that

\[
B_2 = \frac{d\ln Y}{dt} = \left( \frac{1}{Y} \right) \frac{dY}{dt} \\
\frac{dY}{Y} = \frac{\text{relative change in } Y}{\text{absolute change in } t}
\]

**FIGURE 5-3** Semilog model
rate (see footnote 1). In our example the relative change is 0.0107, and hence
the growth rate is 1.07 percent.

Because of this, semilog models like Eq. (5.19) are known as growth mod-
els and such models are routinely used to measure the growth rate of many
variables, whether economic or not.

The interpretation of the intercept term 5.3593 is as follows. From
Eq. (5.15) it is evident that

\[ b_1 = \text{the estimate of } \ln Y_0 = 5.3593 \]

Therefore, if we take the antilog of 5.3593 we obtain

\[ \text{antilog (5.3593)} \approx 212.5761 \]

which is the value of \( Y \) when \( t = 0 \), that is, at the beginning of the period. Since
our sample begins in 1975, we can interpret the value of \( \approx 213 \) (millions) as the
population figure at the end of 1974. But remember the warning given previ-
ously that often the intercept term has no particular physical meaning.

**Instantaneous versus Compound Rate of Growth**

Notice from Eq. (5.16) that

\[ b_2 = \text{the estimate of } B_2 = \ln (1 + r) \]

Therefore,

\[ \text{antilog (} b_2 \text{)} = (1 + r) \]

which means that

\[ r = \text{antilog (} b_2 \text{)} - 1 \quad (5.20) \]

And since \( r \) is the compound rate of growth, once we have obtained \( b_2 \) we can
easily estimate the compound rate of growth of \( Y \) from Equation (5.20). For
Example 5.4, we obtain

\[ r = \text{antilog (0.0107)} - 1 \]

\[ = 1.0108 - 1 = 0.010757 \quad (5.21) \]

That is, over the sample period, the compound rate of growth of the U.S. population
had been at the rate of 1.0757 percent per year.

Earlier we said that the growth rate in \( Y \) was 1.07 percent but now we say it
is 1.0757 percent. What is the difference? The growth rate of 1.07 percent (or,
more generally, the slope coefficient in regressions like Eq. [5.19], multiplied by
100) gives the **instantaneous** (at a point in time) **growth rate**, whereas the
growth rate of 1.0757 percent (or, more generally, that obtained from Equation
[5.20]) is the **compound** (over a period of time) **growth rate**. In the present
example the difference between the two growth rates may not sound important,
but do not forget the power of compounding.

*The Pink Professor*
In practice, one generally quotes the instantaneous growth rate, although the compound growth rate can be easily computed, as just shown.

**The Linear Trend Model**

Sometimes, as a quick and ready method of computation, researchers estimate the following model:

\[ Y_t = B_1 + B_2 t + u_t \]  

That is, regress \( Y \) on time itself, where time is measured chronologically. Such a model is called, appropriately, the **linear trend model**, and the time variable \( t \) is known as the **trend variable**. If the slope coefficient in the preceding model is positive, there is an **upward trend** in \( Y \), whereas if it is negative, there is a **downward trend** in \( Y \).

For the data in Table 5-4, the results of fitting Equation (5.22) are as follows:

\[ \text{USpop}_t = 209.6731 + 2.7570 t \]  

\[ t = (287.4376)(73.6450) \quad r^2 = 0.9943 \]

As these results show, over the sample period the U.S. population had been increasing at the absolute (note, not the relative) rate of 2.757 million per year. Thus, over that period there was an upward trend in the U.S. population. The intercept value here probably represents the base population in the year 1974, which from this model it is about 210 million.

In practice, both the linear trend and growth models have been used extensively. For comparative purposes, however, the growth model is more useful. People are often interested in finding out the relative performance and not the absolute performance of economic measures, such as GDP, money supply, etc.

Incidentally, note that we cannot compare \( r^2 \) values of the two models because the dependent variables in the two models are not the same (but see Problem 5.16). Statistically speaking, both models give fairly good results, judged by the usual \( t \) test of significance.

Recall that for the log-linear, or double-log, model the slope coefficient gives the elasticity of \( Y \) with respect to the relevant explanatory variable. For the growth model and the linear trend models, we can also measure such elasticities. As a matter of fact, once the functional form of the regression model is known, we can compute elasticities from the basic definition of elasticity given in Eq. (5.7). Table 5-11 at the end of this chapter summarizes the elasticity coefficients for the various models we have considered in the chapter.

A cautionary note: The traditional practice of introducing the trend variable \( t \) in models such as (5.18) and (5.22) has recently been questioned by the new generation of time series econometricians. They argue that such a practice may be justifiable only if the error term \( u_t \) in the preceding models is stationary.

---

\(^{17}\)By trend we mean a sustained upward or downward movement in the behavior of a variable.
Although the precise meaning of *stationarity* will be explained in Chapter 12, for now we state that $u_t$ is stationary if its mean value and its variance do not vary systematically over time. In our classical linear regression model we have assumed that $u_t$ has zero mean and constant variance $\sigma^2$. Of course, in an application we will have to check to see if these assumptions are valid. We will discuss this topic later.

### 5.5 THE LIN-LOG MODEL: WHEN THE EXPLANATORY VARIABLE IS LOGARITHMIC

In the previous section we considered the growth model in which the dependent variable was in the log form but the explanatory variable was in the linear form. For descriptive purposes, we can call such a model a log-lin, or growth, model. In this section we consider a model where the dependent variable is in the linear form but the explanatory variable is in the log form. Appropriately, we call this model the lin-log model.

We introduce this model with a concrete example.

**Example 5.5. The Relationship between Expenditure on Services in Relation to Total Personal Consumption Expenditure in 1992 Billions of Dollars, 1975–2006**

Consider the annual data given in Table 5-5 (found on the textbook’s Web site) on consumer expenditure on various categories in relation to total personal consumption expenditure.

Suppose we want to find out how expenditure on services ($Y$) behaves if total personal consumption expenditure ($X$) increases by a certain percentage. Toward that end, suppose we consider the following model:

$$Y_t = B_1 + B_2 \ln X_{2t} + u_t \tag{5.24}$$

In contrast to the log-lin model in Eq. (5.18) where the dependent variable is in log form, the independent variable here is in log form. Before interpreting this model, we present the results based on this model; the results are based on MINITAB.

$$\hat{Y}_t = -12564.8 + 1844.22 \ln X_t$$

$$\text{se} = (916.351) \quad (114.32)$$

$$t = (-13.71) \quad (16.13)$$

$$p = (0.00) \quad (0.00) \quad r^2 = 0.881 \tag{5.25}$$

Interpreted in the usual fashion, the slope coefficient of $\approx 1844$ means that if the log of total personal consumption increases by a unit, the absolute change in the expenditure on personal services is $\approx \$1844$ billion. What does it mean in everyday language? Recall that a change in the log of a number...
is a relative change. Therefore, the slope coefficient in model (5.25) measures

\[ B_2 = \frac{\text{absolute change in } Y}{\text{relative change in } X} \]

\[ = \frac{\Delta Y}{\Delta X/X} \]

where, as before, \( \Delta Y \) and \( \Delta X \) represent (small) changes in \( Y \) and \( X \). Equation (5.26) can be written, equivalently, as

\[ \Delta Y = B_2 \left( \frac{\Delta X}{X} \right) \]

This equation states that the absolute change in \( Y (= \Delta Y) \) is equal to \( B_2 \) times the relative change in \( X \). If the latter is multiplied by 100, then Equation (5.27) 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 \( (B_2) \). Thus, if in an application we find that \( B_2 = 674 \), the absolute change in \( Y \) is \( (0.01)(674) \), or 6.74. Therefore, when regressions like Eq. (5.24) are estimated by OLS, multiply the value of the estimated slope coefficient \( B_2 \) by 0.01, or what amounts to the same thing, divide it by 100.

Returning to our illustrative regression given in Equation (5.25), we then see that if aggregate personal expenditure increases by 1 percent, on the average, expenditure on services increases by \( \approx \$18.44 \) billion. (Note: Divide the estimated slope coefficient by 100.)

Lin-log models like Eq. (5.24) are thus used in situations that study the absolute change in the dependent variable for a percentage change in the independent variable. Needless to say, models like regression (5.24) can have more than one \( X \) variable in the log form. Each partial slope coefficient will then measure the absolute change in the dependent variable for a percentage change in the given \( X \) variable, holding all other \( X \) variables constant.

### 5.6 RECIPROCAL MODELS

Models of the following type are known as reciprocal models:

\[ Y_i = B_1 + B_2 \left( \frac{1}{X_i} \right) + u_i \]

\[ Y = B_1 + B_2 \ln X, \] using calculus it can be shown that \( \frac{dy}{dx} = B_2 \left( \frac{1}{X} \right) \). Therefore, \( B_2 = X \frac{dy}{dx} = \frac{dy}{dx/X} = \text{Eq. (5.26)}. \]
CHAPTER FIVE: FUNCTIONAL FORMS OF REGRESSION MODELS

This model is nonlinear in $X$ because it enters the model *inversely or reciprocally*, but it is a linear regression model because the parameters are linear.\(^{19}\)

The salient feature of this model is that as $X$ increases indefinitely, the term $(1/X)$ approaches zero (Why?) and $Y$ approaches the limiting or asymptotic value of $B_1$. Therefore, models like regression (5.28) have built into them an asymptote or limit value that the dependent variable will take when the value of the $X$ variable increases indefinitely.

Some likely shapes of the curve corresponding to Eq. (5.28) are shown in Figure 5-4.

In Figure 5-4(a) if we let $Y$ stand for the average fixed cost (AFC) of production, that is, the total fixed cost divided by the output, and $X$ for the output, then as economic theory shows, AFC declines continuously as the output increases (because the fixed cost is spread over a larger number of units) and eventually becomes asymptotic at level $B_1$.

An important application of Figure 5-4(b) is the Engel expenditure curve (named after the German statistician Ernst Engel, 1821–1896), which relates a consumer’s expenditure on a commodity to his or her total expenditure or income. If $Y$ denotes expenditure on a commodity and $X$ the total income, then certain commodities have these features: (1) There is some critical or threshold level of income below which the commodity is not purchased (e.g., an automobile). In Figure 5-4(b) this threshold level of income is at the level $-(B_2/B_1)$. (2) There is a satiety level of consumption beyond which the consumer will not go no matter how high the income (even millionaires do not generally own more than two or three cars at a time). This level is nothing but the asymptote $B_1$ shown in Figure 5-4(b). For such commodities, the reciprocal model of this figure is the most appropriate.

One important application of Figure 5-4(c) is the celebrated Phillips curve of macroeconomics. Based on the British data on the percent rate of change of money wages ($Y$) and the unemployment rate ($X$) in percent, Phillips obtained

\[^{19}\text{If we define } X^* = (1/X)\text{, then Equation (5.28) is linear in the parameters as well as the variables } Y \text{ and } X^*.\]

The Pink Professor
a curve similar to Figure 5-4(c).\textsuperscript{20} As this figure shows, there is asymmetry in the response of wage changes to the level of unemployment. Wages rise faster for a unit change in unemployment if the unemployment rate is below $U_N$, which is called the natural rate of unemployment by economists, than they fall for an equivalent change when the unemployment rate is above the natural level, $B_1$ indicating the asymptotic floor for wage change. (See Figure 5-5 later.) This particular feature of the Phillips curve may be due to institutional factors, such as union bargaining power, minimum wages, or unemployment insurance.

Example 5.6. The Phillips Curve for the United States, 1958 to 1969

Because of its historical importance, and to illustrate the reciprocal model, we have obtained data, shown in Table 5-6, on percent change in the index of hourly earnings ($Y$) and the civilian unemployment rate ($X$) for the United States for the years 1958 to 1969.\textsuperscript{21}

Model (5.28) was fitted to the data in Table 5-6, and the results were as follows:

$$
\hat{Y}_t = -0.2594 + 20.5880 \left( \frac{1}{X_t} \right)
$$

$$
t = (-0.2572) \quad (4.3996) \quad r^2 = 0.6594
$$

This regression line is shown in Figure 5-5(a).


\textsuperscript{21}We chose this period because until 1969 the traditional Phillips curve seems to have worked. Since then it has broken down, although many attempts have been made to resuscitate it with varying degrees of success.
As Figure 5-5 shows, the wage floor is $-0.26$ percent, which is not statistically different from zero. (Why?) Therefore, no matter how high the unemployment rate is, the rate of growth of wages will be, at most, zero.

For comparison we present the results of the following linear regression based on the same data (see Figure 5-5[b]):

$$
\hat{Y}_t = 8.0147 - 0.7883X_t \\
(5.30)
$$

Observe these features of the two models. In the linear model (5.30) the slope coefficient is negative, for the higher the unemployment rate is, the lower the rate of growth of earnings will be, ceteris paribus. In the reciprocal model, however, the slope coefficient is positive, which should be the case because the $X$ variable enters inversely (two negatives make one positive). In other words, a positive slope in the reciprocal model is analogous to the negative slope in

*The Pink Professor*
The linear model suggests that as the unemployment rate increases by 1 percentage point, on the average, the percentage point change in the earnings is a constant amount of $-0.79$ no matter at what $X$ we measure it. On the other hand, in the reciprocal model the percentage point rate of change in the earnings is not constant, but rather depends on at what level of $X$ (i.e., the unemployment rate) the change is measured (see Table 5-11). The latter assumption seems economically more plausible. Since the dependent variable in the two models is the same, we can compare the two $r^2$ values. The $r^2$ for the reciprocal model is higher than that for the linear model, suggesting that the former model fits the data better than the latter model.

As this example shows, once we go beyond the LIV/LIP models to those models that are still linear in the parameters but not necessarily so in the variables, we have to exercise considerable care in choosing a suitable model in a given situation. In this choice the theory underlying the phenomenon of interest is often a big help in choosing the appropriate model. There is no denying that model building involves a good dose of theory, some introspection, and considerable hands-on experience. But the latter comes with practice.

Before we leave reciprocal models, we discuss another application of such a model.

**Example 5.7. Advisory Fees Charged for a Mutual Fund**

The data in Table 5-7 relate to the management fees that a leading mutual fund in the United States pays its investment advisers to manage its assets. The fees depend on the net asset value of the fund. As you can see from Figure 5-6, the higher the net asset value of the fund, the lower the advisory fees are.

---

**TABLE 5-7** MANAGEMENT FEE SCHEDULE OF A MUTUAL FUND

<table>
<thead>
<tr>
<th>Fee (%)</th>
<th>Net asset value ($, in billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5200</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5080</td>
<td>5.0</td>
</tr>
<tr>
<td>0.4840</td>
<td>10.0</td>
</tr>
<tr>
<td>0.4600</td>
<td>15.0</td>
</tr>
<tr>
<td>0.4398</td>
<td>20.0</td>
</tr>
<tr>
<td>0.4238</td>
<td>25.0</td>
</tr>
<tr>
<td>0.4115</td>
<td>30.0</td>
</tr>
<tr>
<td>0.4020</td>
<td>35.0</td>
</tr>
<tr>
<td>0.3944</td>
<td>40.0</td>
</tr>
<tr>
<td>0.3880</td>
<td>45.0</td>
</tr>
<tr>
<td>0.3825</td>
<td>55.0</td>
</tr>
<tr>
<td>0.3738</td>
<td>60.0</td>
</tr>
</tbody>
</table>

---

$22$As shown in Table 5-11, for the reciprocal model the slope is $-B_0(1/X^2)$. 

*The Pink Professor*
The graph suggests that the relationship between the two variables is non-linear. Therefore, a model of the following type might be appropriate:

$$\text{Fees} = B_1 + B_2\left(\frac{1}{\text{assets}}\right) + u_i \quad (5.31)$$

Using the data in Table 5-7 and the EViews output in Figure 5-7, we obtained the following regression results:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.420412</td>
<td>0.012858</td>
<td>32.69715</td>
<td>0.0000</td>
</tr>
<tr>
<td>1/assets</td>
<td>0.054930</td>
<td>0.022099</td>
<td>2.485610</td>
<td>0.0322</td>
</tr>
</tbody>
</table>

Dependent Variable: Fees
Method: Least Squares
Sample: 1 12
Included observations: 12

R-squared: 0.381886
Adjusted R-squared: 0.320075
S.E. of regression: 0.041335
Sum squared resid: 0.017086
Mean dependent var: 0.432317
S.D. dependent var: 0.050129
F-statistic: 6.178255
Prob (F-statistic): 0.032232

It is left as an exercise for you to interpret these regression results (see Problem [5.20]).

The Pink Professor
5.7 POLYNOMIAL REGRESSION MODELS

In this section we consider regression models that have found extensive use in applied econometrics relating to production and cost functions. In particular, consider Figure 5-8, which depicts the total cost of production (TC) as a function of output as well as the associated marginal cost (MC) and the average cost (AC) curves.

Letting $Y$ stand for TC and $X$ for the output, mathematically, the total cost function can be expressed as

$$Y_i = B_1 + B_2X_i + B_3X_i^2 + B_4X_i^3$$  \hspace{1cm} (5.32)

which is called a **cubic function**, or, more generally, a **third-degree polynomial** in the variable $X$—the highest power of $X$ represents the degree of the polynomial (three in the present instance).

Notice that in these types of polynomial functions there is only one explanatory variable on the right-hand side, but it appears with various powers, thus making them multiple regression models.\(^{23}\) (Note: We add the error term $u_i$ to make model (5.32) a regression model.)

Although model (5.32) is nonlinear in the variable $X$, it is linear in the parameters, the $B$'s, and is therefore a linear regression model. Thus, models like regression (5.32) can be estimated by the usual OLS routine. The only “worry” about the model is the likely presence of the problem of **collinearity** because the various powered terms of $X$ are functionally related. But this concern is more apparent than real, for the terms $X^2$ and $X^3$ are **nonlinear functions of X** and do not violate the assumption of no perfect collinearity, that is, no perfect linear relationship between variables. In short, polynomial regression models can be estimated in the usual manner and do not present any special estimation problems.

**Example 5.8. Hypothetical Total Cost Function**

To illustrate the polynomial model, consider the hypothetical cost-output data given in Table 5-8.

The OLS regression results based on these data are as follows (see Figure 5-8):

$$\hat{Y}_i = 141.7667 + 63.4776X_i - 12.9615X_i^2 + 0.9396X_i^3$$

se = (6.3753) \hspace{1cm} (4.7786) \hspace{1cm} (0.9857) \hspace{1cm} (0.0591) \hspace{1cm} (5.33)

$$R^2 = 0.9983$$

\(^{23}\)Of course, one can introduce other $X$ variables and their powers, if needed.
If cost curves are to have the U-shaped average and marginal cost curves shown in price theory texts, then the theory suggests that the coefficients in model (5.32) should have these a priori values:

1. $B_1, B_2,$ and $B_4,$ each is greater than zero.
2. $B_3 < 0.$

The regression results given in regression (5.33) clearly are in conformity with these expectations.

As a concrete example of polynomial regression models, consider the following example.

**Example 5.9. Cigarette Smoking and Lung Cancer**

Table 5-9, on the textbook’s Web site, gives data on cigarette smoking and various types of cancer for 43 states and Washington, D.C., for 1960.

---

**TABLE 5-8** HYPOTHETICAL COST-OUTPUT DATA

<table>
<thead>
<tr>
<th>Y($)</th>
<th>193</th>
<th>226</th>
<th>240</th>
<th>244</th>
<th>257</th>
<th>260</th>
<th>274</th>
<th>297</th>
<th>350</th>
<th>420</th>
<th>Total cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>Output</td>
</tr>
</tbody>
</table>

---

**FIGURE 5-8** Cost-output relationship

If cost curves are to have the U-shaped average and marginal cost curves shown in price theory texts, then the theory suggests that the coefficients in model (5.32) should have these a priori values:  

1. $B_1, B_2,$ and $B_4,$ each is greater than zero.
2. $B_3 < 0.$

The regression results given in regression (5.33) clearly are in conformity with these expectations.

As a concrete example of polynomial regression models, consider the following example.

**Example 5.9. Cigarette Smoking and Lung Cancer**

Table 5-9, on the textbook’s Web site, gives data on cigarette smoking and various types of cancer for 43 states and Washington, D.C., for 1960.

---

$\hat{Y}_i = 141.77 + 63.48X_i - 12.96X_i^2 + 0.94X_i^3$  
[Eq. (5.33)]

---

**The Pink Professor**
For now consider the relationship between lung cancer and smoking. To see if smoking has an increasing or decreasing effect on lung cancer, consider the following model:

\[ Y = \text{number of deaths from lung cancer} \quad \text{and} \quad X = \text{the number of cigarettes smoked} \]

The regression results using MINITAB are as shown in Figure 5-9.

These results show that the slope coefficient is positive but the coefficient of the cigarette-squared variable is negative. What this suggests is that cigarette smoking has an adverse impact on lung cancer, but that the adverse impact increases at a diminishing rate.\(^{25}\) All the slope coefficients are statistically significant on the basis of the one-tail \(t\) test. We use the one-tail \(t\) test because medical research has shown that smoking has an adverse impact on lung and other types of cancer. The \(F\) value of 26.56 is also highly significant, for the estimated \(p\) value is practically zero. This would suggest that both variables belong in the model.

### 5.8 REGRESSION THROUGH THE ORIGIN

There are occasions when the regression model assumes the following form, which we illustrate with the two-variable model, although generalization to multiple regression models is straightforward.

\[ Y_i = B_2X_i + u_i \quad \text{(5.35)} \]

\(^{25}\)Neglecting the error term, if you take the derivative of \(Y\) in Equation (5.34) with respect to \(X\), you will obtain \(\frac{dY}{dX} = B_2 + 2B_3X\), which in the present example gives \(1.57 = 2(0.0192)X = 1.57 - 0.0384X\), which shows that the rate of change of lung cancer with respect to cigarette smoking is declining. If the coefficient of the cigsq variable were positive, then the effect of cigarette smoking on lung cancer would be increasing at an increasing rate. Here \(Y\) = incidence of lung cancer and \(X\) is the number of cigarettes smoked.

\[ The Pink Professor \]
In this model the intercept is absent or zero, hence the name **regression through the origin**. We have already come across an example of this in Okun’s law in Eq. (2.22). For Equation (5.35) it can be shown that

\[ b_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \quad (5.36) \]

\[ \text{var} (b_2) = \frac{\sigma^2}{\sum X_i^2} \quad (5.37) \]

\[ \hat{\sigma}^2 = \frac{\sum e_i^2}{n - 1} \quad (5.38) \]

If you compare these formulas with those given for the two-variable model with intercept, given in Equations (2.17), (3.6), and (3.8), you will note several differences. **First**, in the model without the intercept, we use raw sums of squares and cross products, whereas in the intercept-present model, we use mean-adjusted sums of squares and cross products. **Second**, the d.f. in computing \( \hat{\sigma}^2 \) is now \((n - 1)\) rather than \((n - 2)\), since in Eq. (5.35) we have only one unknown. **Third**, the conventionally computed \( r^2 \) formula we have used thus far explicitly assumes that the model has an intercept term. Therefore, you should not use that formula. If you use it, sometimes you will get nonsensical results because the computed \( r^2 \) may turn out to be negative. **Finally**, for the model that includes the intercept, the sum of the estimated residuals, \( \sum \hat{u}_i = \sum e_i \) is always zero, but this need not be the case for a model without the intercept term.

For all these reasons, one may use the zero-intercept model only if there is strong theoretical reason for it, as in Okun’s law or some areas of economics and finance. An example is given in Problem 5.22. For now we will illustrate the zero-intercept model using the data given in Table 2-13, which relates to U.S. real GDP and the unemployment rate for the period 1960 to 2006. Similar to Equation (2.22), we add the variable representing the year and obtain the following results:

\[ \hat{Y}_i = 0.00005 \text{Year} - 3.070X_{t-1} \quad (5.39) \]

\[ t = (2.55) \quad (-2.92) \]

where \( Y = \) change in the unemployment rate in percentage points and Year, \( X_{t-1} = \) percentage growth rate in real GDP from one year prior to the data in \( Y \) and Year.

---

For comparison, we re-estimate Equation (5.39) with the intercept added.

\[ \hat{Y}_t = 3.128 - 0.0015Year - 3.294X_{t-1} \]

\[ t = (3.354)(-0.90) \quad (-3.05) \quad R^2 = 0.182 \]  

(5.40)

As you will notice, the intercept term is significant in Equation (5.40), but now the Year variable is not. Also notice that we have given the $R^2$ value for Eq. (5.40) but not for Eq. (5.39) for reasons stated before.\(^{27}\)

### 5.9 A NOTE ON SCALING AND UNITS OF MEASUREMENT

Variables, economic or not, are expressed in various units of measurement. For example, we can express temperature in Fahrenheit or Celsius. GDP can be measured in millions or billions of dollars. Are regression results sensitive to the unit of measurement? The answer is that some results are and some are not. To show this, consider the data given in Table 5.10.

This table gives data on gross private domestic investment measured in billions of dollars (GDIB), the same data expressed in millions of dollars (GDIM), gross domestic product measured in billions of dollars (GDPB), and the same data expressed in millions of dollars (GDPM). Suppose we want to

<table>
<thead>
<tr>
<th>Year</th>
<th>GDPB</th>
<th>GDPM</th>
<th>GDIB</th>
<th>GDIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>1389.8</td>
<td>1389800</td>
<td>8304.3</td>
<td>8304300</td>
</tr>
<tr>
<td>1998</td>
<td>1509.1</td>
<td>1509100</td>
<td>8747.0</td>
<td>8747000</td>
</tr>
<tr>
<td>1999</td>
<td>1625.7</td>
<td>1625700</td>
<td>9268.4</td>
<td>9268400</td>
</tr>
<tr>
<td>2000</td>
<td>1735.5</td>
<td>1735500</td>
<td>9817.0</td>
<td>9817000</td>
</tr>
<tr>
<td>2001</td>
<td>1614.3</td>
<td>1614300</td>
<td>10128.0</td>
<td>10128000</td>
</tr>
<tr>
<td>2002</td>
<td>1582.1</td>
<td>1582100</td>
<td>10469.6</td>
<td>10469600</td>
</tr>
<tr>
<td>2003</td>
<td>1664.1</td>
<td>1664100</td>
<td>10960.8</td>
<td>10960800</td>
</tr>
<tr>
<td>2004</td>
<td>1888.6</td>
<td>1888600</td>
<td>11685.9</td>
<td>11685900</td>
</tr>
<tr>
<td>2005</td>
<td>2077.2</td>
<td>2077200</td>
<td>12433.9</td>
<td>12433900</td>
</tr>
<tr>
<td>2006</td>
<td>2209.2</td>
<td>2209200</td>
<td>13194.7</td>
<td>13194700</td>
</tr>
</tbody>
</table>

**TABLE 5-10**

GROSS PRIVATE DOMESTIC INVESTMENT AND GROSS DOMESTIC PRODUCT, UNITED STATES, 1997–2006

Variables: GDPB = Gross private domestic product (billions of dollars), GDPM = Gross private domestic product (millions of dollars), GDIB = Gross private domestic investment (billions of dollars), GDIM = Gross private domestic investment (millions of dollars).

\(^{27}\) For Eq. (5.39) we can compute the so-called “raw” $R^2$, which is discussed in Problem 5.23.
find out how GDI behaves in relation to GDP. Toward that end, we estimate the following regression models:

\[
\begin{align*}
\hat{GDI}_t & = 461.511 + 5.8046GDP_B_t \\ 
\text{se} & = (1331.451) (0.762) \\
 t & = (0.3466) (7.6143) \quad r^2 = 0.8787 \\
\end{align*}
\]

\[
\begin{align*}
\hat{GDI}_t & = 461511.076 + 5.8046GDP_M_t \\ 
\text{se} & = (1331451) (0.762) \\
 t & = (0.3466) (7.6143) \quad r^2 = 0.8787 \\
\end{align*}
\]

\[
\begin{align*}
\hat{GDI}_t & = 461.511 + 0.0058GDP_M_t \\ 
\text{se} & = (1331.451) (0.00076) \\
 t & = (0.3466) (7.6143) \quad r^2 = 0.8787 \\
\end{align*}
\]

\[
\begin{align*}
\hat{GDI}_t & = 461511.076 + 5804.626GDP_B_t \\ 
\text{se} & = (1331.451) (0.762) \\
 t & = (0.3466) (7.6143) \quad r^2 = 0.8787 \\
\end{align*}
\]

At first glance these results may look different. But they are not if we take into account the fact that 1 billion is equal to 1,000 million. All we have done in these various regressions is to express variables in different units of measurement. But keep in mind these facts. First, the \( r^2 \) value in all these regressions is the same, which should not be surprising because \( r^2 \) is a pure number, devoid of units in which the dependent variable (\( Y \)) and the independent variable (\( X \)) are measured. Second, the intercept term is always in the units in which the dependent variable is measured; recall that the intercept represents the value of the dependent variable when the independent variable takes the value of zero. Third, when \( Y \) and \( X \) are measured in the same units of measurement the slope coefficients as well as their standard errors remain the same (compare Equations [5.41] and [5.42]), although the intercept values and their standard errors are different. But the \( t \) ratios remain the same. Third, when the \( Y \) and \( X \) variables are measured in different units of measurement, the slope coefficients are different, but the interpretation does not change. Thus, in Equation (5.43) if GDP changes by a million, GDI changes by 0.0058 billions of dollars, which is 5.8 millions of dollars. Likewise, in Equation (5.44) if GDP increases by a billion dollars, GDI increases by 5804.6 millions. All these results are perfectly commonsensical.

5.10 REGRESSION ON STANDARDIZED VARIABLES

We saw in the previous section that the units in which the dependent variable (\( Y \)) and the explanatory variables (the \( X \)'s) are measured affect the interpretation of the regression coefficients. This can be avoided if we express all the variables 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 $Y$ on $X$, if we redefine these variables as

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

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

(5.45)  (5.46)

where $\bar{Y}$ = sample mean of $Y$
$S_Y$ = sample standard deviation of $Y$
$\bar{X}$ = sample mean of $X$
$S_X$ = 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 1.28

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

$$ Y_i = B_1 + B_2 X_i + u_i $$

(5.47)

we could run the regression on the standardized variables as

$$ Y_i^* = B_1^* + B_2^* X_i^* + u_i^* $$

$$ = B_2^* X_i^* + u_i^* $$

(5.48)

since it is easy to show that in the regression involving standardized variables the intercept value is always zero.29 The regression coefficients of the standardized explanatory variables, denoted by starred $B$ coefficients ($B^*$), are known in the literature as the beta coefficients. Incidentally, note that Eq. (5.48) 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, the average value of the (standardized) regressand increases by $B_2^*$ standard deviation units. Thus, unlike the traditional model in Eq. (5.47), we measure the effect not in terms of the original units in which $Y$ and $X$ are measured, but in standard deviation units.

28For proof, see Gujarati and Porter, op.cit., pp. 183–184.
29Recall from Eq. (2.16) that Intercept = Mean value of $Y$ – Slope × Mean value of $X$. But for the standardized variables, the mean value is always zero. This can be easily generalized to more than one $X$ variable.
It should be added that if there is more than one $X$ variable, we can convert each variable into the standardized form. To show this, let us return to the Cobb-Douglas production function data given for real GDP, employment, and real fixed capital for Mexico, 1955–1974, in Table 5-2. The results of fitting the logarithmic function are given in Eq. (5.11). The results of regressing the standardized logs of GDP on standardized employment and standardized fixed capital, using EViews, are as follows:

\[
\begin{align*}
\text{Dependent Variable: SLGDP} \\
\text{Method: Least Squares} \\
\text{Sample: 1955 1974} \\
\text{Included observations: 20}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>$t$-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLE</td>
<td>0.167964</td>
<td>0.089220</td>
<td>1.882590</td>
<td>0.0760</td>
</tr>
<tr>
<td>SLK</td>
<td>0.831995</td>
<td>0.089220</td>
<td>9.325223</td>
<td>0.0000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.995080</td>
<td>Mean dependent var</td>
<td>6.29E-06</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.994807</td>
<td>S.D. dependent var</td>
<td>0.999999</td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.072063</td>
<td>Sum squared resid</td>
<td>0.093475</td>
<td></td>
</tr>
</tbody>
</table>

where $\text{SLGDP} =$ standardized log of GDP
\[
\text{SLE} = \text{standardized log of employment}
\]
\[
\text{SLK} = \text{standardized log of capital}
\]

The interpretation of the regression coefficients is as follows: Holding capital constant, a standard deviation increase in employment increases the GDP, on average, by $\approx 0.17$ standard deviation units. Likewise, holding employment constant, one a standard deviation increase in capital, on average, increases GDP by $\approx 0.83$ standard deviation units. (Note that all variables are in the logarithmic form.) Relatively speaking, capital has more impact on GDP than employment. 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.

Incidentally, we have not introduced the intercept term in the regression results. (Why?) If you include intercept in the model, its value will be almost zero.

### 5.11 SUMMARY OF FUNCTIONAL FORMS

In this chapter we discussed several regression models that, although linear in the parameters, were not necessarily linear in the variables. For each model, we noted its special features and also the circumstances in which it might be
appropriate. In Table 5-11 we summarize the various functional forms that we discussed in terms of a few salient features, such as the slope coefficients and the elasticity coefficients. Although for double-log models the slope and elasticity coefficients are the same, this is not the case for other models. But even for these models, we can compute elasticities from the basic definition given in Eq. (5.7).

As Table 5-11 shows, for the linear-in-variable (LIV) models, the slope coefficient is constant but the elasticity coefficient is variable, whereas for the log-log, or log-linear, model, the elasticity coefficient is constant but the slope coefficient is variable. For other models shown in Table 5-11, both the slope and elasticity coefficients are variable.

### 5.12 SUMMARY

In this chapter we considered models that are linear in parameters, or that can be rendered as such with suitable transformation, but that are not necessarily linear in variables. There are a variety of such models, each having special applications. We considered five major types of nonlinear-in-variable but linear-in-parameter models, namely:

1. The log-linear model, in which both the dependent variable and the explanatory variable are in logarithmic form.
2. The log-lin or growth model, in which the dependent variable is logarithmic but the independent variable is linear.
3. The lin-log model, in which the dependent variable is linear but the independent variable is logarithmic.
4. The reciprocal model, in which the dependent variable is linear but the independent variable is not.\(^{30}\)

\(^{30}\)The dependent variable can also be reciprocal and the independent variable linear, as in Problem 5.15. See also Problem 5.20.
5. The polynomial model, in which the independent variable enters with various powers.

Of course, there is nothing that prevents us from combining the features of one or more of these models. Thus, we can have a multiple regression model in which the dependent variable is in log form and some of the $X$ variables are also in log form, but some are in linear form.

We studied the properties of these various models in terms of their relevance in applied research, their slope coefficients, and their elasticity coefficients. We also showed with several examples the situations in which the various models could be used. Needless to say, we will come across several more examples in the remainder of the text.

In this chapter we also considered the regression-through-the-origin model and discussed some of its features.

It cannot be overemphasized that in choosing among the competing models, the overriding objective should be the economic relevance of the various models and not merely the summary statistics, such as $R^2$. Model building requires a proper balance of theory, availability of the appropriate data, a good understanding of the statistical properties of the various models, and the elusive quality that is called practical judgment. Since the theory underlying a topic of interest is never perfect, there is no such thing as a perfect model. What we hope for is a reasonably good model that will balance all these criteria.

Whatever model is chosen in practice, we have to pay careful attention to the units in which the dependent and independent variables are expressed, for the interpretation of regression coefficients may hinge upon units of measurement.

**KEY TERMS AND CONCEPTS**

The key terms and concepts introduced in this chapter are

- **Double-log, log-linear, or constant elasticity model**
- **Linear vs. log-linear regression model**
  - a) Functional form
  - b) High $r^2$ value criterion
- **Cobb-Douglas (C-D) production function**
  - a) Returns to scale parameter
  - b) Constant returns to scale
  - c) Increasing and decreasing returns to scale
- **Semilog models**
  - a) Instantaneous growth rate
  - b) Compound growth rate
- **Linear trend model**
  - a) Trend variable

- **Log-lin, or growth, model**
- **Lin-log model**
- **Reciprocal models**
  - a) Asymptotic value
  - b) Engel expenditure curve
  - c) The Phillips curve
- **Polynomial regression models**
  - a) Cubic function or third-degree polynomial
  - b) Regression through the origin
- **Scaling and units of measurement**
  - a) Standardized variables
  - b) Beta coefficients

_Philip C. B. Miller_
QUESTIONS

5.1. Explain briefly what is meant by
   a. Log-log model
   b. Log-lin model
   c. Lin-log model
   d. Elasticity coefficient
   e. Elasticity at mean value

5.2. What is meant by a slope coefficient and an elasticity coefficient? What is the
   relationship between the two?

5.3. Fill in the blanks in Table 5-12.

5.4. Complete the following sentences:
   a. In the double-log model the slope coefficient measures . . .
   b. In the lin-log model the slope coefficient measures . . .
   c. In the log-lin model the slope coefficient measures . . .
   d. Elasticity of Y with respect to X is defined as . . .
   e. Price elasticity is defined as . . .
   f. Demand is said to be elastic if the absolute value of the price elasticity is . . ., but demand is said to be inelastic if it is . . .

5.5. State with reason whether the following statements are true (T) or false (F):
   a. For the double-log model, the slope and elasticity coefficients are the same.
   b. For the linear-in-variable (LIV) model, the slope coefficient is constant but the elasticity coefficient is variable, whereas for the log-log model, the
      elasticity coefficient is constant but the slope is variable.
   c. The $R^2$ of a log-log model can be compared with that of a log-lin model but not with that of a lin-log model.
   d. The $R^2$ of a lin-log model can be compared with that of a linear (in variables)
      model but not with that of a double-log or log-lin model.
   e. Model A: $\ln Y = -0.6 + 0.4X$; $r^2 = 0.85$
      Model B: $\hat{Y} = 1.3 + 2.2X$; $r^2 = 0.73$
      Model A is a better model because its $r^2$ is higher.

5.6. The Engel expenditure curve relates a consumer’s expenditure on a commodity to his or her total income. Letting $Y =$ the consumption expenditure on a commodity and $X =$ the consumer income, consider the following models:
   a. $Y_i = B_1 + B_2X_i + u_i$
   b. $Y_i = B_1 + B_2(1/X_i) + u_i$
   c. $\ln Y_i = B_1 + B_2 \ln X_i + u_i$

The Pink Professor
d. \( \ln Y_i = B_1 + B_2 (1/X_i) + u_i \)

e. \( Y_i = B_1 + B_2 \ln X_i + u_i \)

f. \( \ln(Y) = B_1 - B_2 \left( \frac{1}{X} \right) \). This model is known as the log-inverse model.

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

5.7. The growth model Eq. (5.18) was fitted to several U.S. economic time series and the following results were obtained:

<table>
<thead>
<tr>
<th>Time series and period</th>
<th>( B_1 )</th>
<th>( B_2 )</th>
<th>( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real GNP (1954–1987)</td>
<td>7.2492</td>
<td>0.0302</td>
<td>0.9839</td>
</tr>
<tr>
<td>(1982 dollars)</td>
<td>( t = (529.29) )</td>
<td>( (44.318) )</td>
<td></td>
</tr>
<tr>
<td>Labor force participation rate (1973–1987)</td>
<td>4.1056</td>
<td>0.053</td>
<td>0.9464</td>
</tr>
<tr>
<td>S&amp;P 500 index (1954–1987)</td>
<td>3.6960</td>
<td>0.0456</td>
<td>0.8633</td>
</tr>
<tr>
<td>S&amp;P 500 index (1954–1987 quarterly data)</td>
<td>3.7115</td>
<td>0.0114</td>
<td>0.8524</td>
</tr>
</tbody>
</table>

a. In each case find out the instantaneous rate of growth.
b. What is the compound rate of growth in each case?
c. For the S&P data, why is there a difference in the two slope coefficients? How would you reconcile the difference?

PROBLEMS

5.8. Refer to the cubic total cost (TC) function given in Eq. (5.32).

a. The marginal cost (MC) is the change in the TC for a unit change in output; that is, it is the rate of change of the TC with respect to output. (Technically, it is the derivative of the TC with respect to \( X \), the output.) Derive this function from regression (5.32).
b. The average variable cost (AVC) is the total variable cost (TVC) divided by the total output. Derive the AVC function from regression (5.32).
c. The average cost (AC) of production is the TC of production divided by total output. For the function given in regression (5.32), derive the AC function.
d. Plot the various cost curves previously derived and confirm that they resemble the stylized textbook cost curves.

5.9. Are the following models linear in the parameters? If not, is there any way to make them linear-in-parameter (LIP) models?

a. \( Y_i = \frac{1}{B_1 + B_2 X_i} \)

b. \( Y_i = \frac{X_i}{B_1 + B_2 X_i^2} \)

5.10. Based on 11 annual observations, the following regressions were obtained:

Model A: \( \hat{Y}_t = 2.6911 - 0.4795X_t \)  
\[ se = (0.1216) \ (0.1140) \  \ r^2 = 0.6628 \]
Model B: \[ \ln \hat{Y}_t = 0.7774 - 0.2530 \ln X_t \]
\[ \text{se} = (0.0152) (0.0494) \]
\[ r^2 = 0.7448 \]

where \( Y \) = the cups of coffee consumed per person per day and \( X \) = the price of coffee in dollars per pound.

a. Interpret the slope coefficients in the two models.
b. You are told that \( \bar{Y} = 2.43 \) and \( \bar{X} = 1.11 \). At these mean values, estimate the price elasticity for Model A.
c. What is the price elasticity for Model B?
d. From the estimated elasticities, can you say that the demand for coffee is price inelastic?
e. How would you interpret the intercept in Model B? (Hint: Take the antilog.)
f. Since the \( r^2 \) of Model B is larger than that of Model A, Model B is preferable to Model A. Comment on this statement.

5.11. Refer to the Cobb-Douglas production function given in regression (5.11).
a. Interpret the coefficient of the labor input \( X_2 \). Is it statistically different from 1?
b. Interpret the coefficient of the capital input \( X_3 \). Is it statistically different from zero? And from 1?
c. What is the interpretation of the intercept value of \(-1.6524\)?
d. Test the hypothesis that \( B_2 = B_3 = 0 \).

5.12. In their study of the demand for international reserves (i.e., foreign reserve currency such as the dollar or International Monetary Fund [IMF] drawing rights), Mohsen Bahami-Oskooee and Margaret Malixi\(^\text{31}\) obtained the following regression results for a sample of 28 less developed countries (LDC):

\[ \ln(\frac{R}{P}) = 0.1223 + 0.4079 \ln(\frac{Y}{P}) + 0.5040 \ln \sigma_{BP} - 0.0918 \ln \sigma_{EX} \]
\[ t = (2.5128) (17.6377) (15.2437) (-2.7449) \]
\[ R^2 = 0.8268 \]
\[ F = 1151 \]
\[ n = 1120 \]

where \( R \) = the level of nominal reserves in U.S. dollars
\( P \) = U.S. implicit price deflator for GNP
\( Y \) = the nominal GNP in U.S. dollars
\( \sigma_{BP} \) = the variability measure of balance of payments
\( \sigma_{EX} \) = the variability measure of exchange rates

(Notes: The figures in parentheses are \( t \) ratios. This regression was based on quarterly data from 1976 to 1985 (40 quarters) for each of the 28 countries, giving a total sample size of 1120.)
a. A priori, what are the expected signs of the various coefficients? Are the results in accord with these expectations?
b. What is the interpretation of the various partial slope coefficients?


*The Pink Professor*
c. Test the statistical significance of each estimated partial regression coefficient (i.e., the null hypothesis is that individually each true or population regression coefficient is equal to zero).

d. How would you test the hypothesis that all partial slope coefficients are simultaneously zero?

5.13. Based on the U.K. data on annual percentage change in wages ($Y$) and the percent annual unemployment rate ($X$) for the years 1950 to 1966, the following regression results were obtained:

$$\hat{Y}_t = -1.4282 + 8.7243\left(\frac{1}{X_t}\right)$$

$$\text{se} = (2.0675) \quad (2.8478) \quad r^2 = 0.3849$$

$$F(1,15) = 9.39$$

a. What is the interpretation of 8.7243?
b. Test the hypothesis that the estimated slope coefficient is not different from zero. Which test will you use?
c. How would you use the $F$ test to test the preceding hypothesis?
d. Given that $\bar{Y} = 4.8$ percent and $\bar{X} = 1.5$ percent, what is the rate of change of $Y$ at these mean values?
e. What is the elasticity of $Y$ with respect to $X$ at the mean values?
f. How would you test the hypothesis that the true $r^2 = 0$?

5.14. Table 5-13 gives data on the Consumer Price Index, $Y(1980 = 100)$, and the money supply, $X$ (billions of German marks), for Germany for the years 1971 to 1987.

<table>
<thead>
<tr>
<th>Year</th>
<th>$Y$</th>
<th>$X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971</td>
<td>64.1</td>
<td>110.02</td>
</tr>
<tr>
<td>1972</td>
<td>67.7</td>
<td>125.02</td>
</tr>
<tr>
<td>1973</td>
<td>72.4</td>
<td>132.27</td>
</tr>
<tr>
<td>1974</td>
<td>77.5</td>
<td>137.17</td>
</tr>
<tr>
<td>1975</td>
<td>82.0</td>
<td>159.51</td>
</tr>
<tr>
<td>1976</td>
<td>85.6</td>
<td>176.16</td>
</tr>
<tr>
<td>1977</td>
<td>88.7</td>
<td>190.80</td>
</tr>
<tr>
<td>1978</td>
<td>91.1</td>
<td>216.20</td>
</tr>
<tr>
<td>1979</td>
<td>94.9</td>
<td>232.41</td>
</tr>
<tr>
<td>1980</td>
<td>100.0</td>
<td>237.97</td>
</tr>
<tr>
<td>1981</td>
<td>106.3</td>
<td>240.77</td>
</tr>
<tr>
<td>1982</td>
<td>111.9</td>
<td>249.25</td>
</tr>
<tr>
<td>1983</td>
<td>115.6</td>
<td>275.08</td>
</tr>
<tr>
<td>1984</td>
<td>118.4</td>
<td>283.89</td>
</tr>
<tr>
<td>1985</td>
<td>121.0</td>
<td>296.05</td>
</tr>
<tr>
<td>1986</td>
<td>120.7</td>
<td>325.73</td>
</tr>
<tr>
<td>1987</td>
<td>121.1</td>
<td>354.93</td>
</tr>
</tbody>
</table>

170  PART ONE:  THE LINEAR REGRESSION MODEL

a. Regress the following:
   1. Y on X
   2. ln Y on ln X
   3. ln Y on X
   4. Y on ln X
b. Interpret each estimated regression.
c. For each model, find the rate of change of Y with respect to X.
d. For each model, find the elasticity of Y with respect to X. For some of these models, the elasticity is to be computed at the mean values of Y and X.
e. Based on all these regression results, which model would you choose and why?

5.15. Based on the following data, estimate the model:

\[
\frac{1}{Y_i} = B_1 + B_2 X_i + u_i
\]

<table>
<thead>
<tr>
<th>Y</th>
<th>86</th>
<th>79</th>
<th>76</th>
<th>69</th>
<th>65</th>
<th>62</th>
<th>52</th>
<th>51</th>
<th>51</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>3</td>
<td>7</td>
<td>12</td>
<td>17</td>
<td>25</td>
<td>35</td>
<td>45</td>
<td>55</td>
<td>70</td>
<td>120</td>
</tr>
</tbody>
</table>

a. What is the interpretation of \(B_2\)?
b. What is the rate of change of Y with respect to X?
c. What is the elasticity of Y with respect to X?
d. For the same data, run the regression

\[
Y_i = B_1 + B_2 \left( \frac{1}{X_i} \right) + u_i
\]
e. Can you compare the \(r^2\)'s of the two models? Why or why not?
f. How do you decide which is a better model?

5.16. \textit{Comparing two} \(r^2\) \textit{when dependent variables are different.} Suppose you want to compare the \(r^2\) values of the growth model (5.19) with the linear trend model (5.23) of the consumer credit outstanding regressions given in the text. Proceed as follows:
a. Obtain ln \(Y_i\), that is, the estimated log value of each observation from model (5.19).
b. Obtain the antilog values of the values obtained in (a).
c. Compute \(r^2\) between the values obtained in (b) and the actual Y values using the definition of \(r^2\) given in Question 3.5.
d. This \(r^2\) value is comparable with the \(r^2\) value obtained from linear model (5.23).

Use the preceding steps to compare the \(r^2\) values of models (5.19) and (5.23).

5.17. Based on the GNP/money supply data given in Table 5-14 (found on the textbook’s Web site), the following regression results were obtained (\(Y = \) GNP, \(X = M2\)):


\textit{The Pink Professor}
CHAPTER FIVE: FUNCTIONAL FORMS OF REGRESSION MODELS

<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept</th>
<th>Slope</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-linear</td>
<td>0.7826</td>
<td>0.8539</td>
<td>0.997</td>
</tr>
<tr>
<td>$t = 11.40$</td>
<td>$t = 108.93$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-lin</td>
<td>7.2392</td>
<td>0.0001</td>
<td>0.832</td>
</tr>
<tr>
<td>(growth model)</td>
<td>$t = 80.85$</td>
<td>$t = 14.07$</td>
<td></td>
</tr>
<tr>
<td>Lin-log</td>
<td>-24299</td>
<td>3382.4</td>
<td>0.899</td>
</tr>
<tr>
<td></td>
<td>$t = -15.45$</td>
<td>$t = 18.84$</td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>703.28</td>
<td>0.4718</td>
<td>0.991</td>
</tr>
<tr>
<td>(LIV model)</td>
<td>$t = 8.04$</td>
<td>$t = 65.58$</td>
<td></td>
</tr>
</tbody>
</table>

a. For each model, interpret the slope coefficient.

b. For each model, estimate the elasticity of the GNP with respect to money supply and interpret it.

c. Are all $r^2$ values directly comparable? If not, which ones are?

d. Which model will you choose? What criteria did you consider in your choice?

e. According to the monetarists, there is a one-to-one relationship between the rate of changes in the money supply and the GDP. Do the preceding regressions support this view? How would you test this formally?

5.18. Refer to the energy demand data given in Table 5-3. Instead of fitting the log-linear model to the data, fit the following linear model:

$$Y_t = B_1 + B_2 X_{2t} + B_3 X_{3t} + u_t$$

a. Estimate the regression coefficients, their standard errors, and obtain $R^2$ and adjusted $R^2$.

b. Interpret the various regression coefficients.

c. Are the estimated partial regression coefficients individually statistically significant? Use the $p$ values to answer the question.

d. Set up the ANOVA table and test the hypothesis that $B_2 = B_3 = 0$.

e. Compute the income and price elasticities at the mean values of $Y$, $X_2$, and $X_3$. How do these elasticities compare with those given in regression (5.12)?

f. Using the procedure described in Problem 5.16, compare the $R^2$ values of the linear and log-linear regressions. What conclusion do you draw from these computations?

g. Obtain the normal probability plot for the residuals obtained from the linear-in-variable regression above. What conclusions do you draw?

h. Obtain the normal probability plot for the residuals obtained from the log-linear regression (5.12) and decide whether the residuals are approximately normally distributed.

i. If the conclusions in (g) and (h) are different, which regression would you choose and why?

5.19. To explain the behavior of business loan activity at large commercial banks, Bruce J. Summers used the following model:33

$$Y_t = \frac{1}{A + Bt}$$

(A)

where Y is commercial and industrial (C&I) loans in millions of dollars, and $t$ is time, measured in months. The data used in the analysis was collected monthly for the years 1966 to 1967, a total of 24 observations.

For estimation purposes, however, the author used the following model:

$$\frac{1}{Y_t} = A + Bt$$  \hspace{1cm} (B)

The regression results based on this model for banks including New York City banks and excluding New York City banks are given in Equations (1) and (2), respectively:

$$\frac{1}{Y_t} = 52.00 - 0.2t$$  \hspace{1cm} (1)

$$\hat{\frac{1}{Y_t}} = 26.79 - 0.14t$$  \hspace{1cm} (2)

*Durbin-Watson (DW) statistic (see Chapter 10).

**a.** Why did the author use Model (B) rather than Model (A)?

**b.** What are the properties of the two models?

**c.** Interpret the slope coefficients in Models (1) and (2). Are the two slope coefficients statistically significant?

**d.** How would you find out the standard errors of the intercept and slope coefficients in the two regressions?

**e.** Is there a difference in the behavior of New York City and the non–New York City banks in their C&I activity? How would you go about testing the difference, if any, formally?

5.20. Refer to regression (5.31).

**a.** Interpret the slope coefficient.

**b.** Using Table 5-11, compute the elasticity for this model. Is this elasticity constant or variable?

5.21. Refer to the data given in Table 5-5 (found on the textbook’s Web site). Fit an appropriate Engle curve to the various expenditure categories in relation to total personal consumption expenditure and comment on the statistical results.

5.22. Table 5-15 gives data on the annual rate of return $Y \text{ (\%)}$ on Afuture mutual fund and a return on a market portfolio as represented by the Fisher Index, $X \text{ (\%)}$. Now consider the following model, which is known in the finance literature as the characteristic line.

$$Y_t = B_1 + B_2X_t + u_t$$  \hspace{1cm} (1)

*The Pink Professor*
In the literature there is no consensus about the prior value of $B_1$. Some studies have shown it to be positive and statistically significant and some have shown it to be statistically insignificant. In the latter case, Model (1) becomes a regression-through-the-origin model, which can be written as:

$$Y_t = B_2 X_t + u_t$$  \hspace{1cm} (2)

Using the data given in Table 5-15, to estimate both these models and decide which model fits the data better.

5.23. Raw $R^2$ for the regression-through-the-origin model. As noted earlier, for the regression-through-the-origin regression model the conventionally computed $R^2$ may not be meaningful. One suggested alternative for such models is the so-called “raw” $R^2$, which is defined (for the two-variable case) as follows:

$$\text{Raw } r^2 = \frac{(\sum X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$$

If you compare the raw $R^2$ with the traditional $r^2$ computed from Eq. (3.43), you will see that the sums of squares and cross-products in the raw $r^2$ are not mean-corrected.

For model (2) in Problem 5.22 compute the raw $r^2$. Compare this with the $r^2$ value that you obtained for Model (1) in Problem (5.22). What general conclusion do you draw?

5.24. For regression (5.39) compute the raw $r^2$ value and compare it with that given in Eq. (5.40).

5.25. Consider data on the weekly stock prices of Qualcomm, Inc., a digital wireless telecommunications designer and manufacturer, over the time period of 1995 to 2000. The complete data can be found in Table 5-16 on the textbook’s Web site.

**The Pink Professor**
a. Create a scattergram of the closing stock price over time. What kind of pattern is evident in the plot?
b. Estimate a linear model to predict the closing stock price based on time. Does this model seem to fit the data well?
c. Now estimate a squared model by using both time and time-squared. Is this a better fit than in part (b)?
d. Now attempt to fit a cubic or third-degree polynomial to the data as follows:

\[ Y_i = B_0 + B_1 X_i + B_2 X_i^2 + B_3 X_i^3 + u_i \]

where \( Y \) = stock price and \( X \) = time. Which model seems to be the best estimator for the stock prices?

5.26. Table 5-17 on the textbook’s Web site contains data about several magazines. The variables are: magazine name, cost of a full-page ad, circulation (projected, in thousands), percent male among the predicted readership, and median household income of readership. The goal is to predict the advertisement cost.

a. Create scattergrams of the cost variable versus each of the three other variables. What types of relationships do you see?
b. Estimate a linear regression equation with all the variables and create a residuals versus fitted values plot. Does the plot exhibit constant variance from left to right?
c. Now estimate the following mixed model:

\[ \ln Y_i = B_0 + B_1 \ln Circ + B_2 \text{PercMale} + B_3 \text{MedIncome} + u_i \]

and create another residual plot. Does this model fit better than the one in part (b)?

5.27. Refer to Example 4.5 (Table 4-6) about education, GDP, and population for 38 countries.

a. Estimate a linear (LIV) model for the data. What are the resulting equation and relevant output values (i.e., \( F \) statistic, \( t \) values, and \( R^2 \))?
b. Now attempt to estimate a log-linear model (where both of the independent variables are also in the natural log format).
c. With the log-linear model, what does the coefficient of the GDP variable indicate about education? What about the population variable?
d. Which model is more appropriate?

5.28. Table 5-18 on the textbook’s Web site contains data on average life expectancy for 40 countries. It comes from the World Almanac and Book of Facts, 1993, by Pharos Books. The independent variables are the ratio of the number of people per television set and the ratio of number of people per physician.

a. Try fitting a linear (LIV) model to the data. Does this model seem to fit well?
b. Create two scattergrams, one of the natural log of life expectancy versus the natural log of people per television, and one of the natural log of life expectancy versus the natural log of people per physician. Do the graphs appear linear?
c. Estimate the equation for a log-linear model. Does this model fit well?
d. What do the coefficients of the log-linear model indicate about the relationships of the variables to life expectancy? Does this seem reasonable?

5.29 Refer to Example 5.6 in the chapter. It was shown that the percentage change in the index of hourly earnings and the unemployment rate from 1958–1969 followed the traditional Phillips curve model. An updated version of the data, from 1965–2007, can be found in Table 5-19 on the textbook’s Web site.

a. Create a scattergram using the percentage change in hourly earnings as the $Y$ variable and the unemployment rate as the $X$ variable. Does the graph appear linear?

b. Now create a scattergram as above, but use $1/X$ as the independent variable. Does this seem better than the graph in part (a)?

c. Fit Eq. (5.29) to the new data. Does this model seem to fit well? Also create a regular linear (LIV) model as in Eq. (5.30). Which model is better? Why?

APPENDIX 5A: Logarithms

Consider the numbers 5 and 25. We know that

$$25 = 5^2 \quad (5A.1)$$

We say that the exponent 2 is the logarithm of 25 to the base 5. More formally, the logarithm of a number (e.g., 25) to a given base (e.g., 5) is the power (2) to which the base (5) must be raised to obtain the given number (25).

More generally, if

$$Y = b^x \quad (b > 0) \quad (5A.2)$$

then

$$\log_b Y = X \quad (5A.3)$$

In mathematics the function (5A.2) is called an exponential function and (5A.3) is called the logarithmic function. As is clear from Eqs. (5A.2) and (5A.3), one function is the inverse of the other function.

Although any (positive) base can be used, in practice, the two commonly used bases are 10 and the mathematical number $e = 2.71828\ldots$.

Logarithms to base 10 are called common logarithms. Thus,

$$\log_{10} 100 = 2 \quad \log_{10} 30 \approx 1.48$$

That is, in the first case $10^2 = 100$ and in the latter case $30 \approx 10^{1.48}$.

Logarithms to the base $e$ are called natural logarithms. Thus,

$$\log_e 100 \approx 4.6051 \quad \text{and} \quad \log_e 30 \approx 3.4012$$

All these calculations can be done routinely on a hand calculator.

By convention, the logarithm to base 10 is denoted by the letters log and to the base $e$ by $\ln$. Thus, in the preceding example, we can write log 100 or log 30 or ln 100 or ln 30.

The Pink Professor
There is a fixed relationship between the common log and natural log, which is

\[ \ln X = 2.3026 \log X \]  \hspace{1cm} (5A.4)

That is, the natural log of the number \( X \) is equal to 2.3026 times the log of \( X \) to the base 10. Thus,

\[ \ln 30 = 2.3026 \log 30 = 2.3026(1.48) = 3.4012 \text{ (approx.)} \]

as before. Therefore, it does not matter whether one uses common or natural logs. But in mathematics the base that is usually preferred is \( e \), that is, the natural logarithm. Hence, in this book all logs are natural logs, unless stated explicitly. Of course, we can convert the log of a number from one basis to the other using Eq. (5A.4).

Keep in mind that logarithms of negative numbers are not defined. Thus, the log of \((-5)\) or the \(\ln (-5)\) is not defined.

Some properties of logarithms are as follows: If \( A \) and \( B \) are any positive numbers, then it can be shown that:

1. \[ \ln (A \times B) = \ln A + \ln B \]  \hspace{1cm} (5A.5)

That is, the log of the product of two (positive) numbers \( A \) and \( B \) is equal to the sum of their logs.

2. \[ \ln (A/B) = \ln A - \ln B \]  \hspace{1cm} (5A.6)

That is, the log of the ratio of \( A \) to \( B \) is the difference in the logs of \( A \) and \( B \).

3. \[ \ln (A \pm B) \neq \ln A \pm \ln B \]  \hspace{1cm} (5A.7)

That is, the log of the sum or difference of \( A \) and \( B \) is not equal to the sum or difference of their logs.

4. \[ \ln (A^k) = k \ln A \]  \hspace{1cm} (5A.8)

That is, the log of \( A \) raised to power \( k \) is \( k \) times the log of \( A \).

5. \[ \ln e = 1 \]  \hspace{1cm} (5A.9)

That is, the log of \( e \) to itself as a base is 1 (as is the log of 10 to the base 10).

6. \[ \ln 1 = 0 \]  \hspace{1cm} (5A.10)

That is, the natural log of the number 1 is zero (so is the common log of number 1).

7. If \( Y = \ln X \),

\[ \frac{dY}{dX} = \frac{1}{X} \]  \hspace{1cm} (5A.11)
That is, the rate of change (i.e., the derivative) of $Y$ with respect to $X$ is $1$ over $X$.

The exponential and (natural) logarithmic functions are depicted in Figure 5A.1.

Although the number whose log is taken is always positive, the logarithm of that number can be positive as well as negative. It can be easily verified that if

$$0 < Y < 1 \text{ then } \ln Y < 0$$
$$Y = 1 \text{ then } \ln Y = 0$$
$$Y > 1 \text{ then } \ln Y > 0$$

Also note that although the logarithmic curve shown in Figure 5A-1(b) is positively sloping, implying that the larger the number is, the larger its logarithmic value will be, the curve is increasing at a decreasing rate (mathematically, the second derivative of the function is negative). Thus, $\ln(10) = 2.3026$ (approx.) and $\ln(20) = 2.9957$ (approx.). That is, if a number is doubled, its logarithm does not double.

This is why the logarithm transformation is called a nonlinear transformation. This can also be seen from Equation (5A.11), which notes that if $Y = \ln X$, $dY/dX = 1/X$. This means that the slope of the logarithmic function depends on the value of $X$; that is, it is not constant (recall the definition of linearity in the variable).

Logarithms and percentages: Since $d(\ln X)/dX = 1/X$, or $d(\ln X) = dX/X$, for very small changes the change in $\ln X$ is equal to the relative or proportional change in $X$. In practice, if the change in $X$ is reasonably small, the preceding relationship can be written as the change in $\ln X \approx$ to the relative change in $X$, where $\approx$ means approximately.

Thus, for small changes,

$$(\ln X_t - \ln X_{t-1}) \approx \frac{(X_t - X_{t-1})}{X_{t-1}} = \text{ relative change in } X$$
Until now we have only observed models that were linear in parameters as well as in variables. But in life, it'll be hardly the case that way. A dependent variable are linearly related.

The real rate of ROI & GDP is non-linear, i.e., way we have liquidity trap.

A typical AD curve is a non-linear, parabolic function.

So we need to account for various set’s which can exist in between the explanatory variable which vary in linear form.

1. Log-linear or constant elasticity model or Double Log
2. Semi-log model /Log-lin /Average model
3. In log model
4. Reciprocal model (Bivariate set)
5. Polynomial log model (of cost)
6. Log through origin.

Any feature of LIV (Linear in Variables) model was that slope of dependent variable for unit change in independent variable was constant. This might not be the case here.

\[ \log y_i = \log \beta_1 + \beta_2 \log x_i + u_i \]

\[ \ln y_i = \ln \beta_1 - \beta_2 \ln x_i \]
The transformation of $x_i$ does not change the validity of the hypotheses, just the interpretation changes.

While estimating this log-linear model,

$$\hat{y}_i = a_1 + b_2 x_i^*$$

$$(y_i^* = \alpha + \beta x_i + u_i)$$

$$\hat{y}_i^* = \log y_i, x_i^* = \log x_i$$

$a_1, b_2$ are unbiased estimates of $\alpha$ & $\beta$; variance $b_1 = \text{anti-log}(a_1)$ is a biased estimate. [Remember $e = \log p_i^\alpha$]

⇒ See Questions (Problem set)

1. $\widehat{y}_2$


Measuring growth rates: The log-lin model

At times we are interested in finding rate of growth of certain eco. variables like GDP, MS, etc., predicting e.g., population rate of growth. 

$$y_t = y_0 (1 + x)^n$$

: $x$: compounded log of pop.

$$\log y_t = \log y_0 + \log (1 + x)$$

or Pop = $A e^{rt}$

$$\log y_t = \beta_1 + \beta_2 + u_t$$

$$\frac{dy}{y} = \beta \frac{dx}{x} \Rightarrow \text{Rate} = \frac{dy}{dx} = \beta \frac{y}{x}$$

Elasticity: $\frac{dy}{dx} \cdot \frac{x}{y} = \beta \cdot x$

$\beta$: Relative change in $y$

due to unit change in $x$

$$\frac{\beta_2}{\beta} = \text{Relative change in regression}
$$

absolute change in regression

$x$: Multiply $\beta_2$ by 100 to get % growth in $y$ for absolute 1 in $x$. 

Siddharth Rathore
(b) **Lin - Log Model**

\[ y_t = \alpha + \beta \log x_t + u_t \]

\[ \exp y_t = \exp \alpha + \beta \exp x_t + \exp u_t \]

Exp. where \( y_t \)'s are extremely sensitive to changes in \( x_t \)

\[ dy = \beta \frac{dx}{x} \]

**Elasticity**: \[ \frac{dy}{dx} \times \frac{x}{y} = \beta \times \frac{1}{y} \]

**Slope**: \[ \frac{dy}{dx} = \beta \times \frac{1}{x} \]

\( \rightarrow \) inversely prop. to \( x \) & \( y \).

**Interpretation**

\( \alpha \): Food exp = \(-128,19 + 257.27 \log \text{Total exp} \)

\( \beta \): Absolute change in \( y \) caused by a relative change in \( x \) holding other variables constant.

**Engel expenditure**

\[ \text{Food exp} = \alpha + \beta_2 \log \text{Total exp} \]

Total exp devoted to food items relative to \( \text{GNP} \). Total exp. as GNP.

\( \alpha \): Must multiply \( \beta_2 \) with 0.01 for 1x change in \( x \) brings \( \beta_2 \) unit change in \( y \).

---

(III) **Reciprocal Models**

\[ y_t = \beta_1 + \beta_2 \left( \frac{1}{x_t} \right) + u_t \]

\( \beta \): Changes in \( y \) caused by a unit change in the reciprocal of \( x \); holding other variables constant.

\[ \Delta x \rightarrow \infty \Rightarrow y \rightarrow -\infty \]

**Asymptotic**

\[ \Delta x \rightarrow 0 \Rightarrow y \rightarrow \frac{1}{\beta_2} \]

\( \beta_1 > 0, \beta_2 > 0 \)

\( \beta_1 < 0, \beta_2 < 0 \)

**Phillips curve**

\( \beta_1 > 0, \beta_2 > 0 \)

\( \beta_1 < 0, \beta_2 < 0 \)

Rearranging GNP \& CM; the \( x \)'s is not a straight

\( \Delta y \): As per capita GNP increases modestly, there is

a dramatic drop in CM turn sharp turns of \( g \).

\[ CM_i = \beta_1 + \beta_2 \left( \frac{1}{\text{GNP}} \right) \]
**Polynomial Models**

\[ Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + u_i \]

**TC**

\[ TC_i = \alpha + \beta_1 q_i + \beta_2 q_i^2 + \beta_3 q_i^3 + u_i \]

Rel m不合理 TC & Q is not Uniform

"Convex & Concave"

Multi collinearity

\[ \cos (x_i, x_i^2) \] would be high

- G-M asssm's - No perfect Multi collinearity

1. \( \beta_1, \beta_2, \beta_3 \rightarrow BLUE \)

**Note**

Not possible to interpret \( \beta_1, \beta_2, \beta_3 \) separately

**Omitted Variables**

\[ Y_i = b_2 x_i + u_i \] (ORF)

\[ E \left( \sum_{i=1}^{n} (Y_i - b_2 x_i)^2 \right) \]

\[ \frac{\partial E}{\partial b_2} = \sum_{i=1}^{n} x_i (Y_i - b_2 x_i) = 2 \sum_{i=1}^{n} (Y_i - b_2 x_i) x_i \]

\[ b_2 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \] : Substitute

**PRF in m & b**

\[ b_2 = \frac{\sum_{i=1}^{n} x_i (Y_i - b_1 x_i + u_i)}{\sum_{i=1}^{n} x_i^2} = \frac{\beta_2 + \sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \]

\[ E \left( \sum_{i=1}^{n} (b_2 - \beta_2)^2 \right) \]

\[ E \left[ \frac{\sum_{i=1}^{n} x_i^2 u_i^2}{\sum_{i=1}^{n} x_i^2} \right] = \frac{\text{MSE}}{\sum_{i=1}^{n} x_i^2} \]

\[ \sum_{i=1}^{n} x_i u_i = 0 \]

From fundamental eq we have \( E \epsilon \beta_i = 0 \)

- but in this model \( E \epsilon \beta_i \) might not be zero

Siddharth Rathore
Suppose we impose the condition \( \sum E_i = 0 \).

\[
\begin{align*}
\sum Y_i &= b_2 \sum X_i + \sum E_i \\
&= b_2 \sum X_i \\
\Rightarrow b_2 &= \frac{\sum Y_i}{\sum X_i} = \frac{\bar{Y}}{\bar{X}}; \text{ this will not be an unbiased estimator.}
\end{align*}
\]

In reg. through origin we cannot have only \( \sum E_i \).

\( \sum E_i = 0 \)

Hence, \( Y_i = \hat{Y}_i + \hat{E}_i \Rightarrow \bar{Y} = \bar{Y}_i + \bar{\hat{E}} \Rightarrow \bar{Y} \neq \bar{Y}_i \)

Show the condition for which \( R^2 \) in a zero intercept model can be negative (HINT: \( R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum E_i^2}{\sum Y_i^2} \))

\( R \) is not minimized as in an intercept model.

\[
\text{Var} (\hat{E}_i^2) = \frac{\sum E_i^2}{n-1}
\]

Conventional \( R^2 \) may not be appropriate for this model.

\[
R^2 = \frac{\sum \hat{E}_i^2}{\sum Y_i^2}; \text{ this can be large \& meaningless}
\]

Need to exercise great caution before using this model \& use it only if we have a strong a-priori expectation. The previous model has clear adv:

a. Can prove intercept to be statistically insignificant.

b. Avoid specification error.