4. Nonlinear Models and Transformations of Variables

Nonlinear relationships are more plausible than linear ones for many economic processes. In this chapter we will first define what is meant by linear regression analysis and then demonstrate some common methods for extending its use to fit nonlinear relationships. The chapter concludes with a brief outline of the kind of technique used to fit models that cannot be recast in linear form.

4.1 Linearity and nonlinearity

Thus far, when we have used the term 'linear regression analysis', we have not defined exactly what we mean by linearity. It is necessary to do so. Consider the model

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4. \quad (4.1)$$

It is linear in two senses. It is **linear in variables**, because every term consists of a straightforward variable multiplied by a parameter. It is also **linear in parameters**, because every term consists of a straightforward parameter multiplied by a variable.

For the purpose of linear regression analysis, only the second type of linearity is important. Nonlinearity in the variables can always be sidestepped by using appropriate definitions. For example, suppose that the relationship were of the form

$$Y = \beta_1 + \beta_2 X_2^2 + \beta_3 \sqrt{X_3} + \beta_4 \log X_4 + \cdots . \quad (4.2)$$

By defining $Z_2 = X_2^2$, $Z_3 = \sqrt{X_3}$, $Z_4 = \log X_4$, etc., the relationship can be rewritten

$$Y = \beta_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \cdots \quad (4.3)$$

and it is now linear in variables as well as in parameters. This type of transformation is only cosmetic, and you will usually see the regression equation presented with the variables written in their original nonlinear form. This avoids the need for explanation and extra notation.

On the other hand, an equation such as

$$Y = \beta_1 X_2^{\beta_2} \quad (4.4)$$

Table 4.1 Average annual percentage rates of growth of employment, *e*, and real GDP, *g*, 1988–97

	e	*g*	$z = 1/g$		*e*	*g*	$z = 1/g$
Australia	1.68	3.04	0.3289	Korea	2.57	7.73	0.1294
Austria	0.65	2.55	0.3922	Luxembourg	3.02	5.64	0.1773
Belgium	0.34	2.16	0.4630	Netherlands	1.88	2.86	0.3497
Canada	1.17	2.03	0.4926	New Zealand	0.91	2.01	0.4975
Denmark	0.02	2.02	0.4950	Norway	0.36	2.98	0.3356
Finland	–1.06	1.78	0.5618	Portugal	0.33	2.79	0.3584
France	0.28	2.08	0.4808	Spain	0.89	2.60	0.3846
Germany	0.08	2.71	0.3690	Sweden	–0.94	1.17	0.8547
Greece	0.87	2.08	0.4808	Switzerland	0.79	1.15	0.8696
Iceland	–0.13	1.54	0.6494	Turkey	2.02	4.18	0.2392
Ireland	2.16	6.40	0.1563	United Kingdom	0.66	1.97	0.5076
Italy	–0.30	1.68	0.5952	United States	1.53	2.46	0.4065
Japan	1.06	2.81	0.3559				

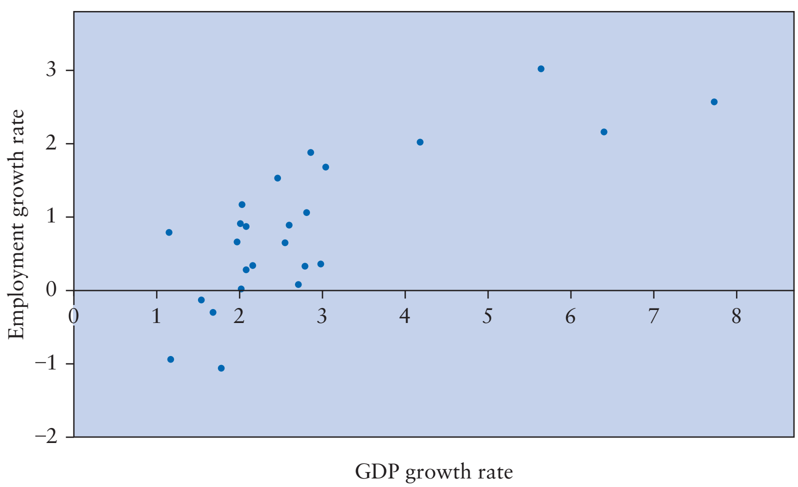

Figure 4.1 Employment and GDP growth rates, 25 OECD countries

is nonlinear in both parameters and variables and cannot be handled by a mere redefinition.

We will begin with an example of a simple model that can be linearized by a cosmetic transformation. Table 4.1 reproduces the data in Exercise 1.4 on average annual rates of growth of employment and GDP for 25 OECD countries. Figure 4.1 plots the data. It is very clear that the relationship is nonlinear. We will consider various nonlinear specifications for the relationship in the course of this chapter, starting with the model

$$e = \beta_1 + \frac{\beta_2}{g} + u. \qquad (4.5)$$

This is nonlinear in g, but if we define $z = 1/g$, we can rewrite the model so that it is linear in variables as well as parameters:

$$e = \beta_1 + \beta_2 z + u. \qquad (4.6)$$

The data for z are given in Table 4.1. In any serious regression application, one would construct z directly from g. The output for a regression of e on z is shown in Table 4.2 and the regression is plotted in Figure 4.2. The regression is shown in equation form as (4.7). The constant term in the regression is an estimate of β_1 and the coefficient of z is an estimate of β_2.

$$\hat{e} = 2.60 - 4.05z. \qquad (4.7)$$

Table 4.2

```
. gen z = 1/g
. reg e z

      Source |       SS           df          MS             Number of obs =      25
-------------+------------------------------------------    F(1, 23)      =   26.06
       Model |   13.1203665        1      13.1203665         Prob > F      =  0.0000
    Residual |   11.5816089       23     .503548214          R-squared     =  0.5311
-------------+------------------------------------------    Adj R-squared =  0.5108
       Total |   24.7019754       24      1.02924898         Root MSE      =  .70961

------------------------------------------------------------------------------------
           e |     Coef.     Std. Err.      t      P>|t|       [95% Conf. Interval]
-------------+----------------------------------------------------------------------
           z |  -4.050817     .793579     -5.10    0.000      -5.69246    -2.409174
       _cons |   2.604753    .3748822      6.95    0.000       1.82925     3.380256
------------------------------------------------------------------------------------
```

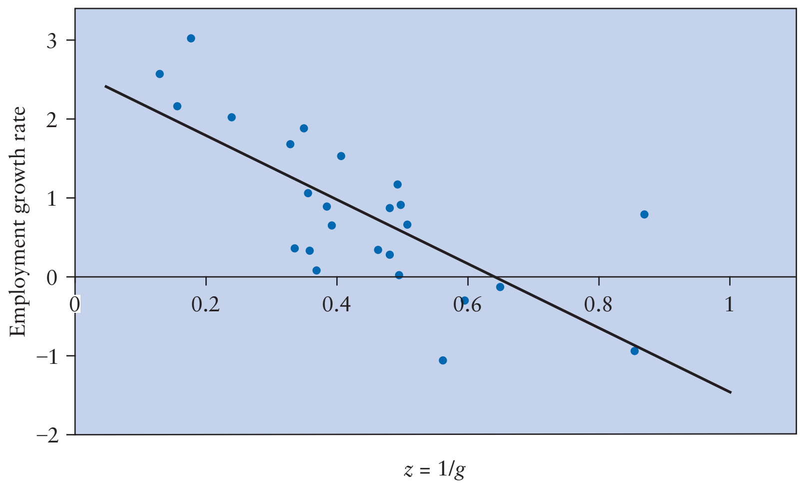

Figure 4.2 Employment growth rate regressed on the reciprocal of GDP growth rate

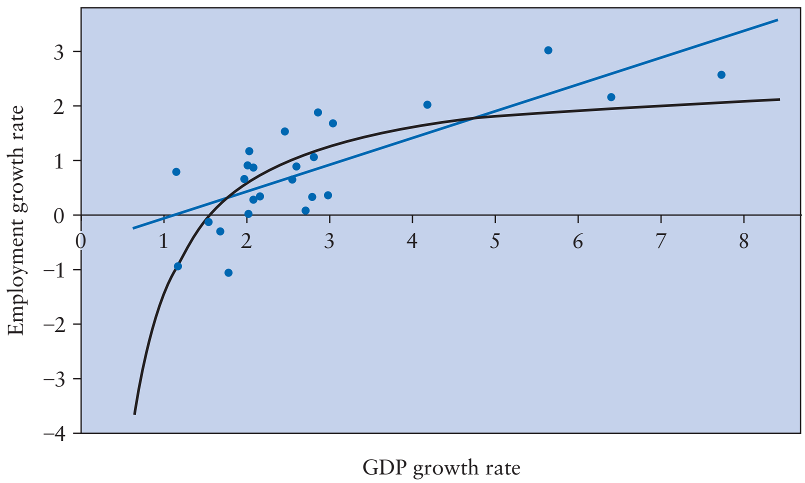

Figure 4.3 Nonlinear and linear regressions of employment growth rate on GDP growth rate

Substituting $z = 1/g$, this becomes

$$\hat{e} = 2.60 - \frac{4.05}{g}. \quad \textbf{(4.8)}$$

Figure 4.3 shows the nonlinear relationship (4.8) plotted in the original diagram. The linear regression reported in Exercise 1.4 is also shown, for comparison.

In this case, it was easy to see that the relationship between e and g was nonlinear. In the case of multiple regression analysis, nonlinearity might be detected using the graphical technique described in Section 3.2.

EXERCISE

4.1

```
. gen Z = 1/SIBLINGS
(11 missing values generated)
. reg S Z

      Source |       SS           df        MS        Number of obs =    529
-------------+--------------------------------------  F(1, 527)     =  30.21
       Model |  169.838682         1   169.838682     Prob > F      = 0.0000
    Residual |   2962.9288       527   5.62225579     R-squared     = 0.0542
-------------+--------------------------------------  Adj R-squared = 0.0524
       Total |  3132.76749       528   5.93327175     Root MSE      = 2.3711

------------------------------------------------------------------------------
           e |     Coef.    Std. Err.     t      P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           z |  2.071194    .3768407    5.50    0.000       1.3309    2.811488
       _cons |   12.7753    .1928491   66.25    0.000     12.39645    13.15415
------------------------------------------------------------------------------
```

It has often been observed that there is a weak tendency for years of schooling to be inversely related to the number of siblings (brothers and sisters) of an individual. The regression shown above has been fitted on the hypothesis that the adverse effect is nonlinear, using *EAEF* Data Set 21. Z is defined as the reciprocal of the number of siblings, for individuals with at least one sibling. Sketch the regression relationship and provide an interpretation of the regression results.

4.2 Logarithmic transformations

Logarithmic models

Next we will tackle functions such as (4.4), which are nonlinear in parameters as well as variables:

$$Y = \beta_1 X^{\beta_2}. \quad (4.9)$$

When you see such a function, you can immediately say that the elasticity of Y with respect to X is constant and equal to β_2. This is easily demonstrated. Regardless of the mathematical relationship connecting Y and X, or the definitions of Y and X, the **elasticity** of Y with respect to X is defined to be the proportional change in Y for a given proportional change in X:

$$\text{elasticity} = \frac{\mathrm{d}Y/Y}{\mathrm{d}X/X}. \quad (4.10)$$

Thus, for example, if Y is demand and X is income, the expression defines the income elasticity of demand for the commodity in question.

The expression may be rewritten

$$\text{elasticity} = \frac{\mathrm{d}Y/\mathrm{d}X}{Y/X}. \quad (4.11)$$

In the case of the demand example, this may be interpreted as the marginal propensity to consume the commodity divided by the average propensity to consume it.

If the relationship between Y and X takes the form (4.9),

$$\frac{\mathrm{d}Y}{\mathrm{d}X} = \beta_1\beta_2 X^{\beta_2 - 1} = \beta_2 \frac{Y}{X}. \quad (4.12)$$

Hence,

$$\text{elasticity} = \frac{\mathrm{d}Y/\mathrm{d}X}{Y/X} = \frac{\beta_2 Y/X}{Y/X} = \beta_2. \quad (4.13)$$

Thus, for example, if you see an Engel curve of the form

$$Y = 0.01X^{0.3}, \quad (4.14)$$

this means that the income elasticity of demand is equal to 0.3. If you are trying to explain this to someone who is not familiar with economic jargon, the easiest way to explain it is to say that a 1 percent change in X (income) will cause a 0.3 percent change in Y (demand).

A function of this type can be converted into a linear equation by means of a **logarithmic transformation**. You will certainly have encountered logarithms in a basic mathematics course. In econometric work they are indispensable. If you are unsure about their use, you should review your notes from that basic math course. The main properties of logarithms are given in Box 4.1.

In the box it is shown that (4.9) may be linearized as

$$\log Y = \log \beta_1 + \beta_2 \log X. \tag{4.15}$$

This is known as a **logarithmic model** or, alternatively, a **loglinear model**, referring to the fact that it is linear in logarithms. If we write $Y' = \log Y$, $Z = \log X$, and $\beta_1' = \log \beta_1$, the equation may be rewritten

$$Y' = \beta_1' + \beta_2 Z. \tag{4.16}$$

The regression procedure is now as follows. First calculate Y' and Z for each observation, taking the logarithms of the original data. Your regression application will almost certainly do this for you, given the appropriate instructions. Second, regress Y' on Z. The coefficient of Z will be a direct estimate of β_2. The constant term will be an estimate of β_1', that is, of $\log \beta_1$. To obtain an estimate of β_1, you have to take the antilog, that is, calculate $\exp(\beta_1')$.

Example: Engel curve

Figure 4.4 plots annual household expenditure on food eaten at home, *FDHO*, and total annual household expenditure, both measured in dollars, for 869

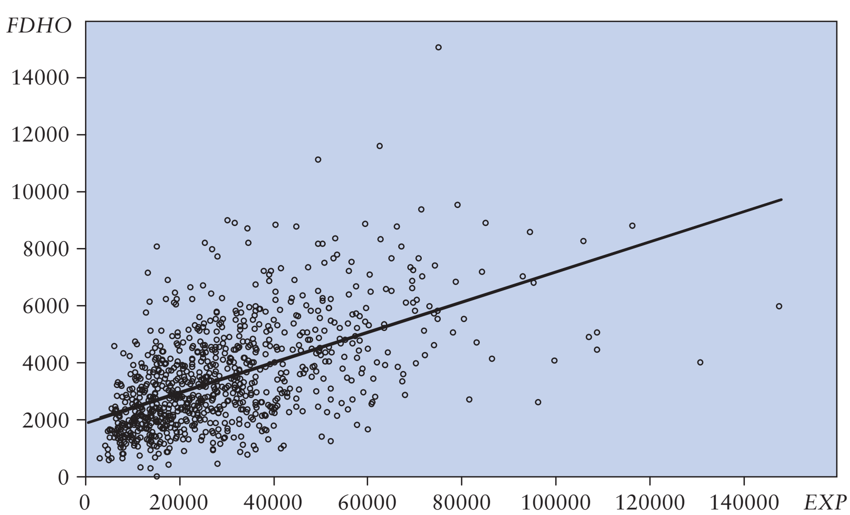

Figure 4.4 Regression of expenditure on food eaten at home on total household expenditure

BOX 4.1 Use of logarithms

First, some basic rules:

1. If $Y = XZ$, $\log Y = \log X + \log Z$
2. If $Y = X/Z$, $\log Y = \log X - \log Z$
3. If $Y = X^n$, $\log Y = n \log X$.

These rules can be combined to transform more complicated expressions. For example, take equation (4.9): if $Y = \beta_1 X^{\beta_2}$,

$$\log Y = \log \beta_1 + \log X^{\beta_2} \qquad \text{using rule 1}$$

$$= \log \beta_1 + \beta_2 \log X \qquad \text{using rule 3.}$$

Thus far, we have not specified whether we are taking logarithms to base e or to base 10. Throughout this text we shall be using e as the base, and so we shall be using what are known as 'natural' logarithms. This is standard in econometrics. Purists sometimes write ln instead of log to emphasize that they are working with natural logarithms, but this is now unnecessary. Nobody uses logarithms to base 10 any more. They were tabulated in the dreaded log tables that were universally employed for multiplying or dividing large numbers until the early 1970s. When the pocket calculator was invented, they became redundant. They are not missed.

With e as base, we can state another rule:

4. If $Y = e^X$, $\log Y = X$.

e^X, also sometimes written $\exp(X)$, is familiarly known as the antilog of X. One can say that $\log e^X$ is the log of the antilog of X, and since log and antilog cancel out, it is not surprising that $\log e^X$ turns out just to be X. Using rule 2 above, $\log e^X = X \log e = X$ since $\log e$ to base e is 1.

representative households in the United States in 1995, the data being taken from the Consumer Expenditure Survey.

When analyzing household expenditure data, it is usual to relate types of expenditure to total household expenditure rather than income, the reason being that the relationship with expenditure tends to be more stable than that with income. The outputs from linear and logarithmic regressions are shown in Tables 4.3 and 4.4.

The linear regression indicates that 5.3 cents out of the marginal dollar are spent on food eaten at home. Interpretation of the intercept is problematic because literally it implies that \$1,916 would be spent on food eaten at home even if total expenditure were zero.

The logarithmic regression, shown in Figure 4.5, indicates that the elasticity of expenditure on food eaten at home with respect to total household expenditure is 0.48. Is this figure plausible? Yes, because food eaten at home is a necessity

Table 4.3

```
. reg FDHO EXP
```

Source	SS	df	MS
Model	915843574	1	915843574
Residual	2.0815e+09	867	2400831.16
Total	2.9974e+09	868	3453184.55

Number of obs	= 869
F(1, 867)	= 381.47
Prob > F	= 0.0000
R-squared	= 0.3055
Adj R-squared	= 0.3047
Root MSE	= 1549.5

FDHO	Coef.	Std. Err.	t	P > |t|	[95% Conf.	Interval]
EXP	.0528427	.0027055	19.531	0.000	.0475325	.0581529
_cons	1916.143	96.54591	19.847	0.000	1726.652	2105.634

Table 4.4

```
. gen LGFDHO = ln(FDHO)
. gen LGEXP = ln(EXP)
. reg LGFDHO LGEXP
```

Source	SS	df	MS
Model	84.4161692	1	84.4161692
Residual	184.579612	866	.213140429
Total	268.995781	867	.310260416

Number of obs	= 868
F(1, 866)	= 396.06
Prob > F	= 0.0000
R-squared	= 0.3138
Adj R-squared	= 0.3130
Root MSE	= .46167

LGFDHO	Coef.	Std. Err.	t	P > |t|	[95% Conf.	Interval]
LGEXP	.4800417	.0241212	19.901	0.000	.4326988	.5273846
_cons	3.166271	.244297	12.961	0.000	2.686787	3.645754

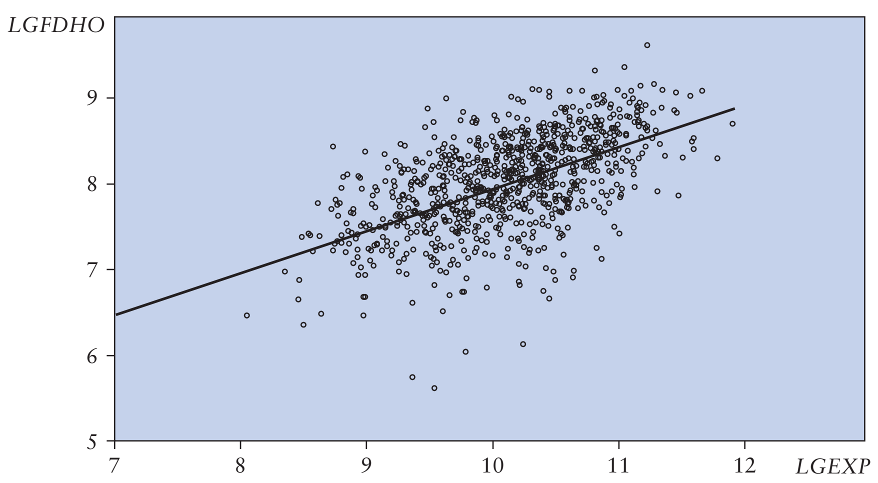

Figure 4.5 Logarithmic regression of expenditure on food eaten at home on total household expenditure

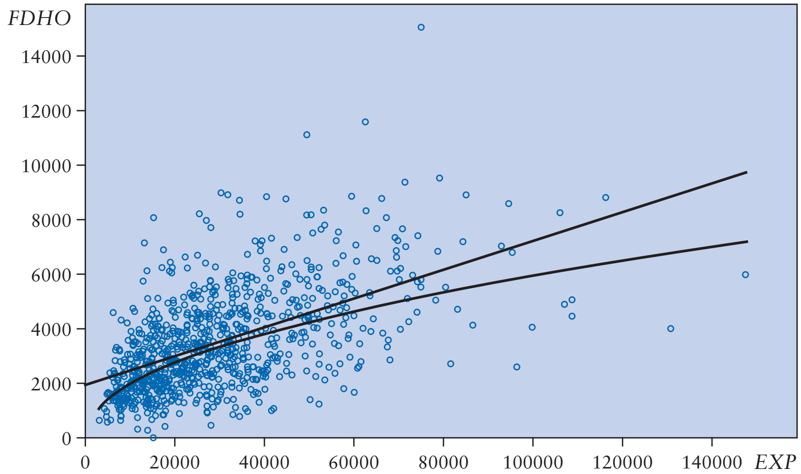

Figure 4.6 Linear and logarithmic regressions of expenditure on food eaten at home on total household expenditure

rather than a luxury, so one would expect the elasticity to be less than 1. The intercept has no economic meaning. Figure 4.6 plots the logarithmic regression line in the original diagram. While there is not much difference between the regression lines over the middle part of the range of observations, it is clear that the logarithmic regression gives a better fit for very low and very high levels of household expenditure.

Semilogarithmic models

Another common functional form is given by equation (4.17):

$$Y = \beta_1 e^{\beta_2 X}. \tag{4.17}$$

Here β_2 should be interpreted as the proportional change in Y per *unit* change in X. Again, this is easily demonstrated. Differentiating,

$$\frac{\mathrm{d}Y}{\mathrm{d}X} = \beta_1 \beta_2 e^{\beta_2 X} = \beta_2 Y. \tag{4.18}$$

Hence,

$$\frac{\mathrm{d}Y/\mathrm{d}X}{Y} = \beta_2 \,. \tag{4.19}$$

In practice, it is often more natural to speak of the percentage change in Y, rather than the proportional change, per unit change in X, in which case one multiplies the estimate of β_2 by 100.

The function can be converted into a model that is linear in parameters by taking the logarithms of both sides:

$$\log Y = \log \beta_1 e^{\beta_2 X} = \log \beta_1 + \log e^{\beta_2 X}$$
$$= \log \beta_1 + \beta_2 X \log e$$
$$= \log \beta_1 + \beta_2 X. \tag{4.20}$$

Note that only the left side is logarithmic in variables, and for this reason (4.20) is described as a **semilogarithmic model**.

The interpretation of β_2 as the proportional change in Y per *unit* change in X is valid only when β_2 is small. When β_2 is large, the interpretation may be a little more complex. Suppose that Y is related to X by (4.17) and that X increases by one unit to X'. Then Y', the new value of Y is given by

$$Y' = \beta_1 e^{\beta_2 X'} = \beta_1 e^{\beta_2 (X+1)}$$
$$= \beta_1 e^{\beta_2 X} e^{\beta_2} = Y e^{\beta_2}$$
$$= Y\left(1 + \beta_2 + \frac{\beta_2^2}{2!} + \dots\right). \tag{4.21}$$

Thus, the proportional change per unit change in X is actually greater than β_2. However, if β_2 is small (say, less than 0.1), β_2^2 and further terms will be very small and can be neglected. In that case, the right side of the equation simplifies to $Y(1 + \beta_2)$ and the original marginal interpretation of β_2 still applies.

Example: semilogarithmic earnings function

For fitting earnings functions, the semilogarithmic model is generally considered to be superior to the linear model. We will start with the simplest possible version:

$$EARNINGS = \beta_1 e^{\beta_2 S}, \tag{4.22}$$

where *EARNINGS* is hourly earnings, measured in dollars, and S is years of schooling. After taking logarithms, the model becomes

$$LGEARN = \beta_1' + \beta_2 S, \tag{4.23}$$

where *LGEARN* is the natural logarithm of *EARNINGS* and β_1' is the logarithm of β_1.

The model was fitted using *EAEF* Data Set 21, with the output shown in Table 4.5. The coefficient of S indicates that every extra year of schooling increases earnings by a proportion 0.110, that is, 11.0 percent, as a first approximation. Strictly speaking, a whole extra year of schooling is not marginal, so it would be more accurate to calculate $e^{0.110}$, which is 1.116. Thus, a more accurate interpretation is that an extra year of schooling raises earnings by 11.6 percent.

Table 4.5

```
. reg LGEARN S
    Source  |        SS          df         MS           Number of obs =     540
------------+------------------------------------        F(1, 538)     = 140.05
     Model  |   38.5643833        1      38.5643833      Prob > F      = 0.0000
  Residual  |    148.14326      538      .275359219      R-squared     = 0.2065
------------+------------------------------------        Adj R-squared = 0.2051
     Total  |   186.707643      539       .34639637      Root MSE      = .52475
--------------------------------------------------------------------------------
    LGEARN  |       Coef.    Std. Err.      t     P > |t|     [95% Conf. Interval]
------------+-------------------------------------------------------------------
         S  |    .1096934    .0092691    11.83    0.000     .0914853     .1279014
    _ cons  |    1.292241    .1287252    10.04    0.000     1.039376     1.545107
--------------------------------------------------------------------------------
```

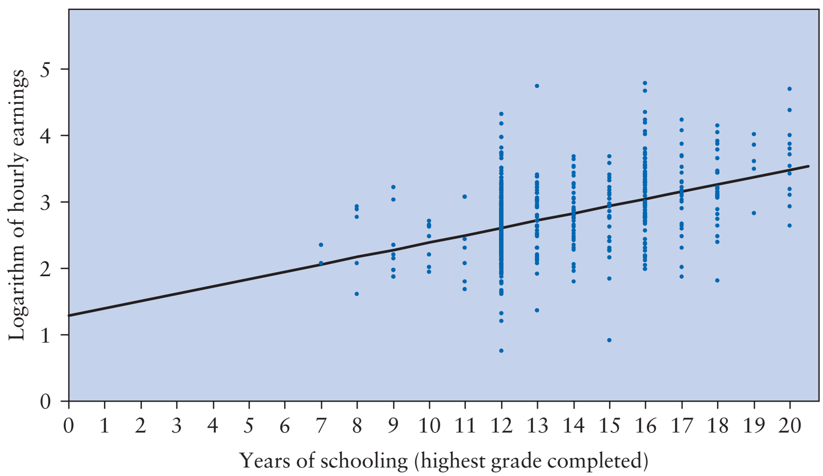

Figure 4.7 Semilogarithmic regression of earnings on schooling

The scatter diagram for the semilogarithmic regression is shown in Figure 4.7. For the purpose of comparison, it is plotted together with the linear regression in a plot with the untransformed variables in Figure 4.8. The two regression lines do not differ greatly in their overall fit, but the semilogarithmic specification has the advantages of not predicting negative earnings for individuals with low levels of schooling and of allowing the increase in earnings per year of schooling to increase with schooling.

The disturbance term

Thus far, nothing has been said about how the disturbance term is affected by these transformations. Indeed, in the discussion above it has been left out altogether.

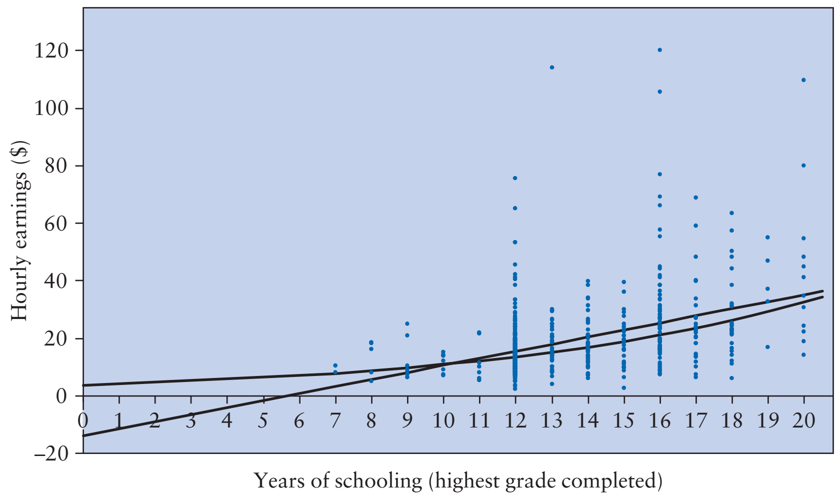

Figure 4.8 Linear and semilogarithmic regressions of earnings on schooling

The fundamental requirement is that the disturbance term should appear in the transformed equation as an additive term (+ u) that satisfies the regression model conditions. If it does not, the least squares regression coefficients will not have the usual properties, and the tests will be invalid.

For example, it is highly desirable that (4.6) should be of the form

$$e = \beta_1 + \beta_2 z + u \tag{4.24}$$

when we take the random effect into account. Working backwards, this implies that the original (untransformed) equation should be of the form

$$e = \beta_1 + \frac{\beta_2}{g} + u. \tag{4.25}$$

In this particular case, if it is true that in the original equation the disturbance term is additive and satisfies the regression model conditions, it will also be true in the transformed equation. No problem here.

What happens when we start off with a model such as

$$Y = \beta_1 X_2^{\beta_2}? \tag{4.26}$$

As we have seen, the regression model, after linearization by taking logarithms, is

$$\log Y = \log \beta_1 + \beta_2 \log X + u \tag{4.27}$$

when the disturbance term is included. Working back to the original equation, this implies that (4.26) should be rewritten

$$Y = \beta_1 X_2^{\beta_2} v, \tag{4.28}$$

where v and u are related by $\log v = u$. Hence, to obtain an additive disturbance term in the regression equation for this model, we must start with a multiplicative disturbance term in the original equation.

The disturbance term v modifies $\beta_1 X_2^{\beta_2}$ by increasing it or reducing it by a random *proportion*, rather than by a random amount. Note that $u = 0$ when $\log v = 0$, which occurs when $v = 1$. The random factor will be zero in the estimating equation (4.27) if v happens to be equal to 1. This makes sense, since if v is equal to 1 it is not modifying $\beta_1 X_2^{\beta_2}$ at all.

For the t tests and the F tests to be valid, u must be normally distributed. This means that $\log v$ must be normally distributed, which will occur only if v is lognormally distributed.

What would happen if we assumed that the disturbance term in the original equation was additive, instead of multiplicative?

$$Y = \beta_1 X_2^{\beta_2} + u. \tag{4.29}$$

The answer is that when you take logarithms, there is no mathematical way of simplifying $\log\left(\beta_1 X_2^{\beta_2} + u\right)$. The transformation does not lead to a linearization. You would have to use a nonlinear regression technique, for example, of the type discussed in the next section.

Example

The central limit theorem suggests that the disturbance term should have a normal distribution. It can be demonstrated that if the disturbance term has a normal distribution, so also will the residuals, provided that the regression equation is correctly specified. An examination of the distribution of the residuals thus provides indirect evidence of the adequacy of the specification of a regression model. Figure 4.9 shows the residuals from linear and semilogarithmic regressions of *EARNINGS* on *S* using *EAEF* Data Set 21, standardized so that they

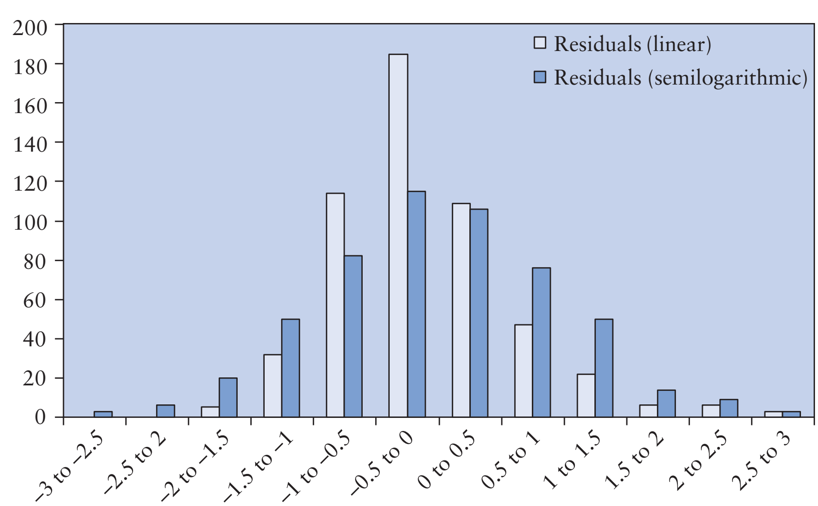

Figure 4.9 Standardized residuals from earnings function regressions

have standard deviation equal to 1, for comparison. The distribution of the residuals from the linear specification is right skewed, while that for the residuals from the semilogarithmic specification is much closer to a normal distribution. This suggests that the semilogarithmic specification is preferable.

Comparing linear and logarithmic specifications

The possibility of fitting nonlinear models, either by means of a linearizing transformation or by the use of a nonlinear regression algorithm, greatly increases the flexibility of regression analysis, but it also makes model specification more complex. You have to ask yourself whether you should start off with a linear relationship or a nonlinear one, and if the latter, what kind.

A graphical inspection of the scatter plot may be sufficient to establish that a relationship is nonlinear. In the case of multiple regression models, the Frisch–Waugh–Lovell technique described in Section 3.2 may be useful. However, it then often happens that several different nonlinear forms might approximately fit the observations if they lie on a curve.

When considering alternative models with the same specification of the dependent variable, the selection procedure is straightforward. The most sensible thing to do is to run regressions based on alternative plausible functions and choose the function that explains the greatest proportion of the variation in the dependent variable. If two or more functions are more or less equally good, you should present the results of each.

However, when alternative models employ different functional forms for the dependent variable, the problem of model selection becomes more complicated because you cannot make direct comparisons of R^2 or the sum of the squares of the residuals. In particular—and this is the most common example of the problem—you cannot compare these statistics for linear and logarithmic dependent variable specifications.

For example, in Section 1.4, the linear regression of earnings on schooling has an R^2 of 0.173, and *RSS* is 92,689. For the semilogarithmic version in Table 4.5, the corresponding figures are 0.207 and 148. *RSS* is much smaller for the logarithmic version, but this means nothing at all. The values of *LGEARN* are much smaller than those of *EARNINGS*, so it is hardly surprising that the residuals are also much smaller. Admittedly, R^2 is unit-free, but it is referring to different concepts in the two equations. In one equation it is measuring the proportion of the variation in earnings explained by the regression, and in the other it is measuring the proportion of the variation in the logarithm of earnings explained. If R^2 is much greater for one model than for the other, you would probably be justified in selecting it without further fuss. But if R^2 is similar for the two models, simple eyeballing will not do.

One procedure under these circumstances, based on Box and Cox (1964), is to scale the observations on Y so that the residual sums of squares in the linear and logarithmic models are rendered directly comparable. The procedure has the following steps:

1. You calculate the geometric mean of the values of Y in the sample. This is equal to the exponential of the mean of log Y, so it is easy to calculate:

$$e^{\frac{1}{n}\sum \log Y_i} = e^{\frac{1}{n}\log(Y_1 \times \cdots \times Y_n)} = e^{\log(Y_1 \times \cdots \times Y_n)^{\frac{1}{n}}} = (Y_1 \times \cdots \times Y_n)^{\frac{1}{n}}. \tag{4.30}$$

2. You scale the observations on Y by dividing by this figure. So

$$Y_i^* = Y_i \,/\, \text{geometric mean of } Y, \tag{4.31}$$

   where Y^* is the scaled value in observation i.
3. You then regress the linear model using Y^* instead of Y as the dependent variable, and the logarithmic model using log Y^* instead of log Y, but otherwise leaving the models unchanged. The residual sums of squares of the two regressions are now comparable, and the model with the lower sum provides the better fit.

Note that the scaled regressions are solely for deciding which model you prefer. You should *not* pay any attention to their coefficients, only to their residual sums of squares. You obtain the coefficients by fitting the unscaled version of the preferred model.

Example

The comparison will be made for the alternative specifications of the earnings function. The mean value of *LGEARN* is 2.7920. The scaling factor is therefore exp(2.7920) = 16.3135. Table 4.6 begins with commands for generating *EARNSTAR*, the scaled version of *EARNINGS*, and its logarithm, *LGEARNST*.

Table 4.6

```
. gen EARNSTAR = EARNINGS/16.3135
. gen LGEARNST = ln(EARNSTAR)
. reg EARNSTAR S EXP
     Source   |        SS          df           MS           Number of obs =    540
--------------+------------------------------------------    F(2, 537)     =  67.54
      Model   |    84.5963381       2       42.298169        Prob > F      = 0.0000
   Residual   |    336.288615     537      .626235783        R-squared     = 0.2010
--------------+------------------------------------------    Adj R-squared = 0.1980
      Total   |    420.884953     539      .780862622        Root MSE      = .79135
-------------------------------------------------------------------------------------
   EARNSTAR   |      Coef.      Std. Err.      t      P > |t|     [95% Conf. Interval]
--------------+----------------------------------------------------------------------
          S   |    .1641662     .0143225    11.46     0.000      .1360312    .1923011
        EXP   |    .0344765     .0078777     4.38     0.000      .0190015    .0499515
      _cons   |   -1.623503     .2619003    -6.20     0.000     -2.137977   -1.109028
-------------------------------------------------------------------------------------
```

Table 4.7

```
. reg LGEARNST S EXP

    Source  |        SS           df          MS          Number of obs =    540
------------+------------------------------------         F(2, 537)     = 100.86
     Model  |   50.9842589         2      25.4921295      Prob > F      = 0.0000
  Residual  |    135.72339       537      .252743742      R-squared     = 0.2731
------------+------------------------------------         Adj R-squared = 0.1980
     Total  |   186.707649       539      .346396379      Root MSE      = .50274

------------------------------------------------------------------------------
  LGEARNST  |      Coef.     Std. Err.     t     P>|t|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
         S  |    .1235911    .0090989    13.58   0.000      .1057173    .141465
       EXP  |    .0350826    .0050046     7.01   0.000      .0252515   .0449137
     _cons  |   -2.282673    .1663823   -13.72   0.000     -2.609513  -1.955833
------------------------------------------------------------------------------
```

EARNSTAR is then regressed on *S* and *EXP*. The residual sum of squares is 336.29. The corresponding regression of *LGEARNST* on *S* and *EXP* follows in Table 4.7. The residual sum of squares is 135.72. Hence, in this case, the semi-logarithmic specification appears to provide the better fit.

EXERCISES

4.2

```
. gen LGS = ln(S)
. gen LGSM = ln(SM)
(4 missing values generated)
. reg LGS LGSM

    Source  |        SS           df          MS          Number of obs =    536
------------+------------------------------------         F(1, 534)     =  56.99
     Model  |   1.62650898         1      1.62650898      Prob > F      = 0.0000
  Residual  |   15.2402109       534      .028539721      R-squared     = 0.0964
------------+------------------------------------         Adj R-squared = 0.0947
     Total  |   16.8667198       535      .031526579      Root MSE      = .16894

------------------------------------------------------------------------------
       LGS  |    Coef.      Std. Err.     t     P>|t|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
      LGSM  |   .200682     .0265831    7.55   0.000     .1484618      .2529022
     _cons  |   2.11373     .0648636   32.59   0.000     1.986311      2.241149
------------------------------------------------------------------------------
```

The output shows the result of regressing *LGS*, the logarithm of years of schooling, on *LGSM*, the logarithm of mother's years of schooling, using *EAEF* Data Set 21. Provide an interpretation of the coefficients and evaluate the regression results.

4.3

```
. reg LGS SM

  Source   |       SS          df          MS            Number of obs =    540
-----------+------------------------------------         F(1, 538)     =  78.13
   Model   |   2.14395934       1      2.14395934        Prob > F      = 0.0000
Residual   |   14.7640299     538      .027442435        R-squared     = 0.1268
-----------+------------------------------------         Adj R-squared = 0.1252
   Total   |   16.9079893     539      .031369182        Root MSE      = .16566
-------------------------------------------------------------------------------
     LGS   |     Coef.     Std. Err.     t      P>|t|    [95% Conf.   Interval]
-------------------------------------------------------------------------------
      SM   |   .0223929    .0025335    8.84     0.000    .0174162     .0273696
   _ cons  |   2.340421    .0301902   77.52     0.000    2.281116     2.399727
-------------------------------------------------------------------------------
```

The output shows the result of regressing *LGS*, the logarithm of years of schooling, on *SM*, mother's years of schooling, using *EAEF* Data Set 21. Provide an interpretation of the coefficients and evaluate the regression results.

4.4 Download the *CES* data set from the website (see Appendix B) and fit linear and logarithmic regressions for your commodity on *EXP*, total household expenditure, excluding observations with zero expenditure on your commodity. Interpret the regression results and perform appropriate tests.

4.5 Repeat the logarithmic regression in Exercise 4.4, adding the logarithm of the size of the household as an additional explanatory variable. Interpret the regression results and perform appropriate tests.

4.6 What is the relationship between weight and height? Using your *EAEF* data set, regress the logarithm of *WEIGHT85* on the logarithm of *HEIGHT*. Interpret the regression results and perform appropriate tests.

4.7 Suppose that the logarithm of *Y* is regressed on the logarithm of *X*, the fitted regression being

$$\widehat{\log Y} = b_1 + b_2 \log X.$$

Suppose $Y^* = \lambda Y$, where λ is a constant, and suppose that $\log Y^*$ is regressed on $\log X$. Determine how the regression coefficients are related to those of the original regression. Determine also how the *t* statistic for b_2 and R^2 for the equation are related to those in the original regression.

4.8* Suppose that the logarithm of *Y* is regressed on the logarithm of *X*, the fitted regression being

$$\widehat{\log Y} = b_1 + b_2 \log X.$$

Suppose $X^* = \mu X$, where λ is a constant, and suppose that $\log Y^*$ is regressed on $\log X^*$. Determine how the regression coefficients are related to those of the original regression. Determine also how the *t* statistic for b_2 and R^2 for the equation are related to those in the original regression.

4.9 Using your *EAEF* data set, regress the logarithm of earnings on *S* and *EXP*. Interpret the regression results and perform appropriate tests.

4.10 Using your *EAEF* data set, evaluate whether the dependent variable of an earnings function should be linear or logarithmic. Calculate the geometric mean of *EARNINGS* by taking the exponential of the mean of *LGEARN*. Define

EARNSTAR by dividing *EARNINGS* by this quantity and calculate *LGEARNST* as its logarithm. Regress *EARNSTAR* and *LGEARNST* on *S* and *EXP* and compare the residual sums of squares.

4.11 Evaluate whether a linear or logarithmic specification of the dependent variable is preferable for the expenditure function for your commodity in the *CES* data set. *Note*: Drop households reporting no expenditure on your commodity.

4.3 Models with quadratic and interactive variables

We come now to models with quadratic terms, such as

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u \tag{4.32}$$

and models with interactive terms, such as

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_2 X_3 + u. \tag{4.33}$$

Of course, the quadratic model may be viewed as a special case of the interactive model with $X_3 = X_2$, but it is convenient to treat it separately. These models can be fitted using OLS with no modification. However, the interpretation of their coefficients has to be approached with care. The usual interpretation of a parameter, that it represents the effect of a unit change in its associated variable, holding all other variables constant, cannot be applied. In the case of the quadratic model, it is not possible for X_2 to change without X_2^2 also changing. In the case of the interactive model, it is not possible for X_2 to change without $X_2 X_3$ also changing, if X_3 is kept constant.

Quadratic variables

By differentiating (4.32), one obtains the change in Y per unit change in X_2:

$$\frac{dY}{dX_2} = \beta_2 + 2\beta_3 X_2 . \tag{4.34}$$

Viewed this way, it can be seen that the impact of a unit change in X_2 on Y, $(\beta_2 + 2\beta_3 X_2)$, changes with X_2. This means that β_2 has an interpretation that is different from that in the ordinary linear model

$$Y = \beta_1 + \beta_2 X_2 + u, \tag{4.35}$$

where it is the unqualified effect of a unit change in X_2 on Y. In (4.34), β_2 should be interpreted as the effect of a unit change in X_2 on Y for the special case where $X_2 = 0$. For nonzero values of X_2, the coefficient will be different.

β_3 also has a special interpretation. If we rewrite the model as

$$Y = \beta_1 + (\beta_2 + \beta_3 X_2) X_2 + u, \tag{4.36}$$

β_3 can be interpreted as the rate of change of the coefficient of X_2, per unit change in X_2.

Only β_1 has a conventional interpretation. As usual, it is the value of Y (apart from the random component) when $X_2 = 0$.

There is a further problem. We have already seen that the estimate of the intercept may have no sensible meaning if $X_2 = 0$ is outside the data range. For example, in the case of the linear regression of earnings on schooling reproduced in Figure 4.10, the intercept is negative, implying that an individual with no schooling would have hourly earnings of –\$13.93. If $X_2 = 0$ lies outside the data range, the same type of distortion can happen with the estimate of β_2.

We will illustrate this with the earnings function. Table 4.8 gives the output of a quadratic regression of earnings on schooling (*SSQ* is defined as the square

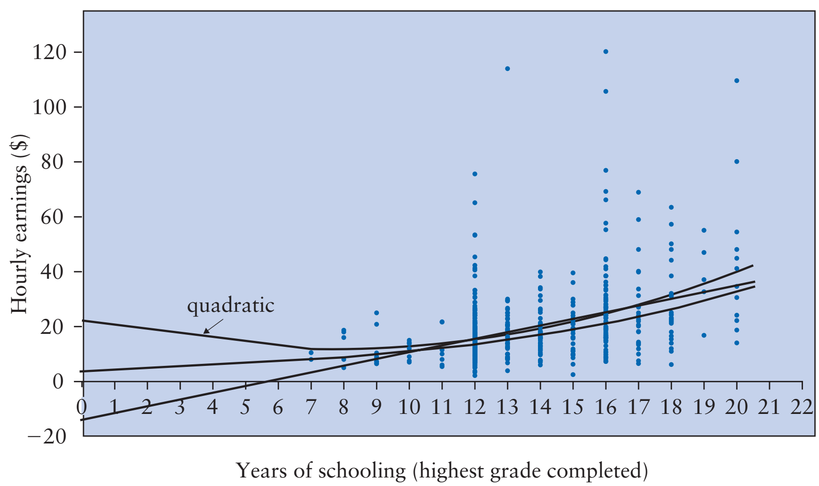

Figure 4.10 Quadratic, linear, and semilogarithmic regressions of earnings on schooling

Table 4.8

```
. gen SSQ = S*S
. reg EARNINGS S SSQ

    Source   |        SS            df          MS            Number of obs =    540
-------------------------------------------------------------  F(2, 537)     =  59.69
     Model   |   20372.4953          2      10186.2477         Prob > F      = 0.0000
  Residual   |   91637.7357        537     170.0647553         R-squared     = 0.1819
-------------------------------------------------------------  Adj R-squared = 0.1788
     Total   |   112010.231        539      207.811189         Root MSE      = 13.063

------------------------------------------------------------------------------------
  EARNINGS   |      Coef.     Std. Err.     t      P > |t|     [95% Conf. Interval]
------------------------------------------------------------------------------------
         S   |   -2.772317    2.119128   -1.31    0.191     -6.935114    1.390481
       SSQ   |    .1829731    .0737308    2.48    0.013      .0381369    .3278092
     _ cons  |    22.25089    14.92883    1.49    0.137     -7.075176    51.57695
------------------------------------------------------------------------------------
```

of schooling). The coefficient of S implies that, for an individual with no schooling, the impact of a year of schooling is to *decrease* hourly earnings by $2.77. The intercept also has no sensible interpretation. Literally, it implies that an individual with no schooling would have hourly earnings of $22.25, which is implausibly high.

The quadratic relationship is illustrated in Figure 4.10. Over the range of the actual data, it fits the observations tolerably well. The fit is not dramatically different from those of the linear and semilogarithmic specifications. However, when one extrapolates beyond the data range, the quadratic function increases as schooling decreases, giving rise to implausible estimates of both β_1 and β_2 for $S = 0$. In this example, we would prefer the semilogarithmic specification, as do all wage-equation studies. The slope coefficient of the semilogarithmic specification has a simple interpretation and the specification does not give rise to predictions outside the data range that are obviously nonsensical.

The data on employment growth rate, e, and GDP growth rate, g, for 25 OECD countries in Exercise 1.4 provide a less problematic example of the use of a quadratic function. *gsq* has been defined as the square of g. Table 4.9 shows the output from the quadratic regression. In Figure 4.11, the quadratic regression is compared with that obtained in Section 4.1. The quadratic specification appears to be an improvement on the hyperbolic function fitted in Section 4.1. It is more satisfactory than the latter for low values of g, in that it does not yield implausibly large negative predicted values of e. The only defect is that it predicts that the fitted value of e starts to fall when g exceeds 7.

Higher-order polynomials

Why stop at a quadratic? Why not consider a cubic, or quartic, or a polynomial of even higher order? There are usually several good reasons for not doing so. Diminishing marginal effects are standard in economic theory, justifying

Table 4.9

```
. gen gsq = g*g
. reg e g gsq
```

Source	SS	df	MS
Model	15.9784642	2	7.98923212
Residual	8.7235112	22	.396523236
Total	24.7019754	24	1.02924898

Number of obs	= 25
F (2, 22)	= 20.15
Prob > F	= 0.0000
R-squared	= 0.6468
Adj R-squared	= 0.6147
Root MSE	= .6297

e	Coef.	Std. Err.	t	P > |t|	[95% Conf.	Interval]
g	1.200205	.3862226	3.11	0.005	.3992287	2.001182
gsq	-.0838408	.0445693	-1.88	0.073	-.1762719	.0085903
_cons	-1.678113	.6556641	-2.56	0.018	-3.037877	-.3183494

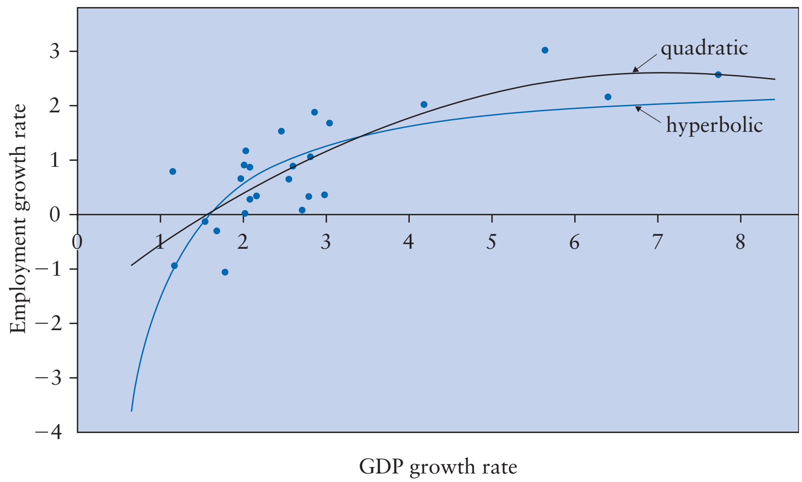

Figure 4.11 Hyperbolic and quadratic regressions of employment growth rate on GDP growth rate

quadratic specifications, at least as an approximation, but economic theory seldom suggests that a relationship might sensibly be represented by a cubic or higher-order polynomial. The second reason follows from the first. There will be an improvement in fit as higher-order terms are added, but because these terms are not theoretically justified, the improvement will be sample-specific. Third, unless the sample is very small, the fits of higher-order polynomials are unlikely to be very different from those of a quadratic over the main part of the data range.

These points are illustrated by Figure 4.12, which shows cubic and quartic regressions with the original linear and quadratic regressions. Over the main

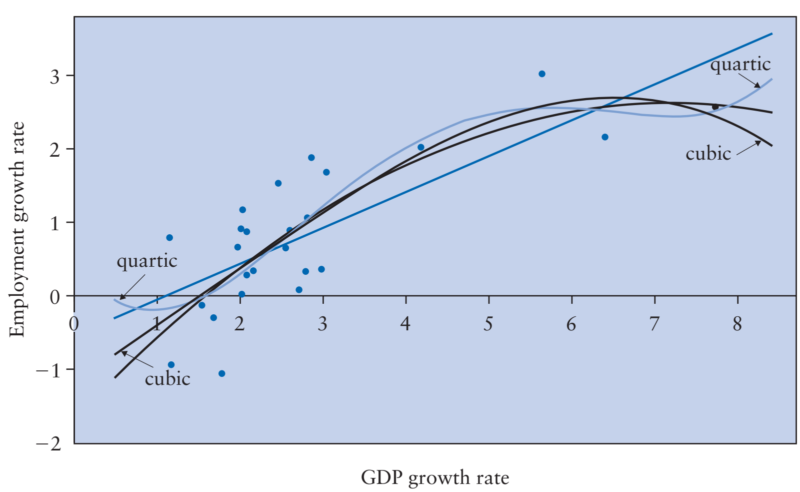

Figure 4.12 Cubic and quartic regressions of employment growth rate on GDP growth rate

data range, from $g = 1.5$ to $g = 4$, the fits of the cubic and quartic are very similar to that of the quadratic. R^2 for the linear specification is 0.590. For the quadratic it improves to 0.647. For the cubic and quartic it is 0.651 and 0.658, relatively small further improvements. Further, the cubic and quartic curves both exhibit implausible characteristics. The cubic declines even more rapidly than the quadratic for high values of g, and the quartic has strange twists at its extremities.

Interactive explanatory variables

We next turn to models with interactive terms, such as

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_2 X_3 + u. \tag{4.37}$$

This is linear in parameters and so may be fitted using straightforward OLS. However, the fact that it is nonlinear in variables complicates the interpretation of the parameters. It is not possible to interpret β_2 as the effect of X_2 on Y, holding X_3 and X_2X_3 constant, because it is not possible to hold both X_3 and X_2X_3 constant if X_2 changes.

To give a proper interpretation to the coefficients, we can rewrite the model as

$$Y = \beta_1 + (\beta_2 + \beta_4 X_3) X_2 + \beta_3 X_3 + u. \tag{4.38}$$

This representation makes explicit the fact that $(\beta_2 + \beta_4 X_3)$, the marginal effect of X_2 on Y, depends on the value of X_3. From this it can be seen that the interpretation of β_2 has a special interpretation. It gives the marginal effect of X_2 on Y, when $X_3 = 0$.

One may alternatively rewrite the model as

$$Y = \beta_1 + \beta_2 X_2 + (\beta_3 + \beta_4 X_2) X_3 + u. \tag{4.39}$$

From this it may be seen that the marginal effect of X_3 on Y, holding X_2 constant, is $(\beta_3 + \beta_4 X_2)$ and that β_3 may be interpreted as the marginal effect of X_3 on Y, when $X_2 = 0$.

If $X_3 = 0$ is a long way outside the range of X_3 in the sample, the interpretation of the estimate of β_2 as an estimate of the marginal effect of X_2 when $X_3 = 0$ should be treated with caution. Sometimes the estimate will be completely implausible, in the same way as the estimate of the intercept in a regression is often implausible if given a literal interpretation. We have just encountered a similar problem with the interpretation of β_2 in the quadratic specification. Often it is of interest to compare the estimates of the effects of X_2 and X_3 on Y in models excluding and including the interactive term, and the changes in the meanings of β_2 and β_3 caused by the inclusion of the interactive term can make such comparisons difficult.

One way of mitigating the problem is to rescale X_2 and X_3 so that they are measured from their sample means:

$$X_2^* = X_2 - \bar{X}_2 \tag{4.40}$$

$$X_3^* = X_3 - \bar{X}_3. \tag{4.41}$$

Substituting for X_2 and X_3, the model becomes

$$\begin{aligned} Y &= \beta_1 + \beta_2\left(X_2^* + \bar{X}_2\right) + \beta_3\left(X_3^* + \bar{X}_3\right) + \beta_4\left(X_2^* + \bar{X}_2\right)\left(X_3^* + \bar{X}_3\right) + u \\ &= \beta_1^* + \beta_2^* X_2^* + \beta_3^* X_3^* + \beta_4 X_2^* X_3^* + u \end{aligned} \tag{4.42}$$

where $\beta_1^* = \beta_1 + \beta_2\bar{X}_2 + \beta_3\bar{X}_3 + \beta_4\bar{X}_2\bar{X}_3$, $\beta_2^* = \beta_2 + \beta_4\bar{X}_3$, and $\beta_3^* = \beta_3 + \beta_4\bar{X}_2$. The point of doing this is that the coefficients of X_2^* and X_3^* now give the marginal effects of the variables when the other variable is held at its sample mean, which is to some extent a representative value. For example, rewriting the new equation as

$$Y = \beta_1^* + \left(\beta_2^* + \beta_4 X_3^*\right)X_2^* + \beta_3^* X_3^* + u \tag{4.43}$$

it can be seen that β_2^* gives the marginal effect of X_2^*, and hence X_2, when $X_3^* = 0$, that is, when X_3 is at its sample mean. β_3^* has a similar interpretation.

Example

Table 4.10 shows the results of regressing the logarithm of hourly earnings on years of schooling and years of work experience for males using *EAEF* Data Set 21. It implies that an extra year of schooling increases earnings by 13.0 percent and that an extra year of work experience increases them by 3.2 percent. In Table 4.11, the interactive variable *SEXP* is defined as the product of *S* and *EXP* and added to the specification. The schooling coefficient now jumps to 23.7 percent, an extraordinarily high figure. But of course it has now changed its meaning. It now estimates the impact of an extra year of schooling

Table 4.10

```
. reg LGEARN S EXP

   Source   |        SS            df            MS            Number of obs =    270
------------+------------------------------------------        F(2, 267)     =  50.41
    Model   |    25.4256872         2        12.7128436        Prob > F      = 0.0000
 Residual   |     7.3402828       267        .252210797        R-squared     = 0.2741
------------+------------------------------------------        Adj R-squared = 0.2686
    Total   |      92.76597       269        .344854907        Root MSE      = .50221
-------------------------------------------------------------------------------------
   LGEARN   |        Coef.     Std. Err.      t       P > |t|    [95% Conf.  Interval]
------------+------------------------------------------------------------------------
        S   |      .1303979     .0129905    10.04     0.000      .1048211    .1559747
      EXP   |      .0321614     .0083783     3.84     0.000      .0156655    .0486573
    _ cons  |      .5969745     .2768371     2.16     0.032      .0519132    1.142036
-------------------------------------------------------------------------------------
```

Table 4.11

```
. gen SEXP = S*EXP
. reg LGEARN S EXP SEXP

    Source   |       SS          df          MS         Number of obs =    270
-------------+--------------------------------------    F(3, 266)     =  35.58
     Model   |  26.5654376        3      8.85514586     Prob > F      = 0.0000
  Residual   |  66.2005325      266      .248874182     R-squared     = 0.2864
-------------+--------------------------------------    Adj R-squared = 0.2783
     Total   |    92.76597      269      .344854907     Root MSE      = .49887

------------------------------------------------------------------------------
    LGEARN   |      Coef.    Std. Err.     t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         S   |    .2371066   .0515064    4.60    0.000     .1356944    .3385187
       EXP   |    .1226418   .0430918    2.85    0.005     .0377974    .2074863
      SEXP   |   -.0065695   .0030699   -2.14    0.033    -.0126139   -.0005252
     _cons   |   -.9003565   .7517877   -1.20    0.232    -2.380568     .579855
------------------------------------------------------------------------------
```

Table 4.12

```
. egen MEANS = mean(S)
. gen S1 = S - MEANS
. egen MEANEXP = mean(EXP)
. gen EXP1 = EXP - MEANEXP
. gen SEXP1 = S1*EXP1

. reg LGEARN S1 EXP1

    Source   |       SS          df          MS         Number of obs =    270
-------------+--------------------------------------    F(2, 267)     =  50.41
     Model   |  25.4256872        2      12.7128436     Prob > F      = 0.0000
  Residual   |  67.3402828      267      .252210797     R-squared     = 0.2741
-------------+--------------------------------------    Adj R-squared = 0.2686
     Total   |    92.76597      269      .344854907     Root MSE      = .50221

------------------------------------------------------------------------------
    LGEARN   |      Coef.    Std. Err.     t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        S1   |    .1303979   .0129905   10.04    0.000     .1048211    .1559747
      EXP1   |    .0321614   .0083783    3.84    0.000     .0156655    .0486573
     _cons   |    2.961112   .0305633   96.88    0.000     2.900936    3.021288
------------------------------------------------------------------------------
```

for those individuals who have no work experience. The experience coefficient has also risen sharply. Now it indicates that an extra year increases earnings by a wholly implausible 12.3 percent. But this figure refers to individuals with no schooling, and every individual in the sample had at least 7 years.

To deal with these problems, we define *S1*, *EXP1*, and *SEXP1* as the corresponding schooling, experience, and interactive variables with the means subtracted, and repeat the regressions. We will refer to the original regressions excluding and including the interactive term, with the output shown in Tables 4.10 and 4.11, as Regressions (1) and (2), and the new ones, with the output shown in Tables 4.12 and 4.13, as Regressions (3) and (4).

Table 4.13

```
. reg LGEARN S1 EXP1 SEXP1
    Source   |         SS             df          MS           Number of obs =    270
-------------+----------------------------------------------   F(3, 266)     =  35.58
     Model   |     26.5654377         3       8.85514589       Prob > F      = 0.0000
  Residual   |     66.2005324       266       .248874182       R-squared     = 0.2864
-------------+----------------------------------------------   Adj R-squared = 0.2783
     Total   |       92.76597       269       .344854907       Root MSE      = .49887
-------------+-----------------------------------------------------------------------
    LGEARN   |      Coef.       Std. Err.     t       P>|t|      [95% Conf. Interval]
-------------+-----------------------------------------------------------------------
        S1   |    .1196959      .0138394     8.65     0.000      .0924473    .1469446
      EXP1   |    .0324933      .0083241     3.90     0.000      .0161038    .0488829
     SEXP1   |   -.0065695      .0030699    -2.14     0.033     -.0126139   -.0005252
     _cons   |    2.933994      .0328989    89.18     0.000      2.869218    2.998769
-------------+-----------------------------------------------------------------------
```

Regression (3) is virtually identical to Regression (1). In particular, comparing Tables 4.10 and 4.12, the slope coefficients are the same, as are the standard errors, t statistics, and R^2. The only difference is in the intercept. This now refers to the logarithm of hourly earnings of an individual with mean schooling and mean experience. It implies an hourly rate of $e^{2.96} = 19.32$ dollars, and since it is in the middle of the data range it is perhaps more informative than the intercept in Regression (1), which suggested that the hourly earnings of an individual with no schooling and no work experience would be $e^{0.60} = 1.82$ dollars.

Regressions (2) and (4) also have much in common. The analysis of variance and goodness of fit statistics are the same, and the results relating to the interactive effect are the same. The only differences are in the output relating to the schooling and work experience slope coefficients, which in Regression (4) now relate to an individual with mean schooling and experience. A comparison of Regressions (3) and (4) allows a more meaningful evaluation of the impact of including the interactive term. We see that, for an individual at the mean, it has little effect on the value of work experience, but it suggests that the value of a year of schooling was overestimated by a small amount in Regression (3). The interactive effect itself suggests that the value of education diminishes for all individuals with increasing experience.

Ramsey's RESET test of functional misspecification

Adding quadratic terms of the explanatory variables and interactive terms to the specification is one way of investigating the possibility that the dependent variable in a model may be a nonlinear function of them. However, if there are many explanatory variables in the model, before devoting time to experimentation with quadratic and interactive terms, it may be useful to have some means of detecting whether there is any evidence of nonlinearity in the first place.

Ramsey's RESET test of functional misspecification is intended to provide a simple indicator.

To implement it, one runs the regression in the original form and then saves the fitted values of the dependent variable, which we will denote $\hat{Y}$. Since, by definition,

$$\hat{Y} = b_1 + \sum_{j=2}^{k} b_j X_j, \quad (4.44)$$

$\hat{Y}^2$ is a linear combination of the squares of the X variables and their interactions. If $\hat{Y}^2$ is added to the regression specification, it should pick up quadratic and interactive nonlinearity, if present, without necessarily being highly correlated with any of the X variables and consuming only one degree of freedom. If the t statistic for the coefficient of $\hat{Y}^2$ is significant, this indicates that some kind of nonlinearity may be present.

Of course the test does not indicate the actual form of the nonlinearity and it may fail to detect other types of nonlinearity. However, it does have the virtue of being very easy to implement.

In principle, one could also include higher powers of $\hat{Y}$. However, the consensus appears to be that this is not usually worthwhile.

EXERCISES

4.12

```
. gen SMSQ = SM*SM
. reg S SM SMSQ

  Source  |       SS         df        MS          Number of obs =    540
----------+--------------------------------------  F(2, 537)     =  51.90
   Model  |  519.131914       2   259.565957       Prob > F      = 0.0000
Residual  |  2685.85142     534   5.00158551       R-squared     = 0.1620
----------+--------------------------------------  Adj R-squared = 0.1589
   Total  |  3204.98333     539   5.94616574       Root MSE      = 2.2364
-------------------------------------------------------------------------
       S  |      Coef.   Std. Err.     t     P>|t|    [95% Conf. Interval]
----------+--------------------------------------------------------------
      SM  |  -.2564658   .1318583   -1.95    0.052   -.5154872   .0025556
    SMSQ  |   .0271172   .0060632    4.47    0.000    .0152068   .0390277
   _cons  |   12.79121   .7366358   17.36    0.000    11.34416   14.23825
-------------------------------------------------------------------------
```

The output shows the result of regression of S on SM and its square, $SMSQ$. Evaluate the regression results. In particular, explain why the coefficient of SM is negative.

4.13 Using your *EAEF* data set, perform a regression parallel to that in Exercise 4.12 and evaluate the results. Define a new variable $SM12$ as $SM - 12$. $SM12$ may be interpreted as the number of years of schooling of the mother after completing high school, if positive, and the number of years of schooling lost before completing high school, if negative. Regress S on $SM12$ and its square, and compare the results with those in your original regression.

4.14*

```
. reg LGS LGSM LGSMSQ

  Source   |       SS          df          MS          Number of obs =    536
-----------+------------------------------------       F(1, 534)     =  56.99
   Model   |    1.62650898       1       1.62650898    Prob > F      = 0.0000
Residual   |    15.2402109     534      .028539721    R-squared     = 0.0964
-----------+------------------------------------       Adj R-squared = 0.0947
   Total   |    16.8667198     535      .031526579    Root MSE      = .16894

-----------------------------------------------------------------------------
     LGS   |     Coef.     Std. Err.     t      P>|t|     [95% Conf. Interval]
-----------+-----------------------------------------------------------------
    LGSM   |   (omitted)
  LGSMSQ   |    .100341    .0132915    7.55    0.000     .0742309    .1264511
   _cons   |    2.11373    .0648636   32.59    0.000     1.986311    2.241149
-----------------------------------------------------------------------------
```

The output shows the results of regressing *LGS*, the logarithm of *S*, on *LGSM*, the logarithm of *SM*, and *LGSMSQ*, the logarithm of *SMSQ*. Explain the regression results.

4.15 Perform a RESET test of functional misspecification. Using your *EAEF* data set, regress *WEIGHT02* on *HEIGHT*. Save the fitted values as *YHAT* and define *YHATSQ* as its square. Add *YHATSQ* to the regression specification and test its coefficient.

4.4 Nonlinear regression

Suppose you believe that a variable Y depends on a variable X according to the relationship

$$Y = \beta_1 + \beta_2 X^{\beta_3} + u, \tag{4.45}$$

and you wish to obtain estimates of β_1, β_2, and β_3 given data on Y and X. There is no way of transforming (4.45) to obtain a linear relationship, and so it is not possible to apply the usual regression procedure.

Nevertheless, one can still use the principle of minimizing the sum of the squares of the residuals to obtain estimates of the parameters. We will describe a simple **nonlinear regression algorithm** that uses the principle. It consists of a series of repeated steps:

1. You start by guessing plausible values for the parameters.
2. You calculate the predicted values of Y from the data on X, using these values of the parameters.
3. You calculate the residual for each observation in the sample, and hence *RSS*, the sum of the squares of the residuals.
4. You then make small changes in one or more of your estimates of the parameters.
5. You calculate the new predicted values of Y, residuals, and *RSS*.
6. If *RSS* is smaller than before, your new estimates of the parameters are better than the old ones and you take them as your new starting point.

7. You repeat steps 4, 5, and 6 again and again until you are unable to make any changes in the estimates of the parameters that would reduce *RSS*.
8. You conclude that you have minimized *RSS*, and you can describe the final estimates of the parameters as the least squares estimates.

Example

We will return to the relationship between employment growth rate, *e*, and GDP growth rate, *g*, in Section 4.1, where *e* and *g* are hypothesized to be related by

$$e = \beta_1 + \frac{\beta_2}{g} + u. \tag{4.46}$$

According to this specification, as *g* becomes large, *e* will tend to a limit of β_1. Looking at the scatter diagram for *e* and *g*, we see that the maximum value of *e* is about 2. So we will take this as our initial value for b_1. We then hunt for the optimal value of b_2, conditional on this guess for b_1. Figure 4.13 shows *RSS* plotted as a function of b_2, conditional on $b_1 = 2$. From this we see that the optimal value of b_2, conditional on $b_1 = 2$, is -2.86.

Next, holding b_2 at -2.86, we look to improve on our guess for b_1. Figure 4.14 shows *RSS* as a function of b_1, conditional on $b_2 = -2.86$. We see that the optimal value of b_1 is 2.08.

We continue to do this until both parameter estimates cease to change. We will then have reached the values that yield minimum *RSS*. These must be the values from the linear regression shown in Table 4.2: $b_1 = 2.60$ and $b_2 = -4.05$. They have been determined by the same criterion, the minimization of *RSS*. All that we have done is to use a different method.

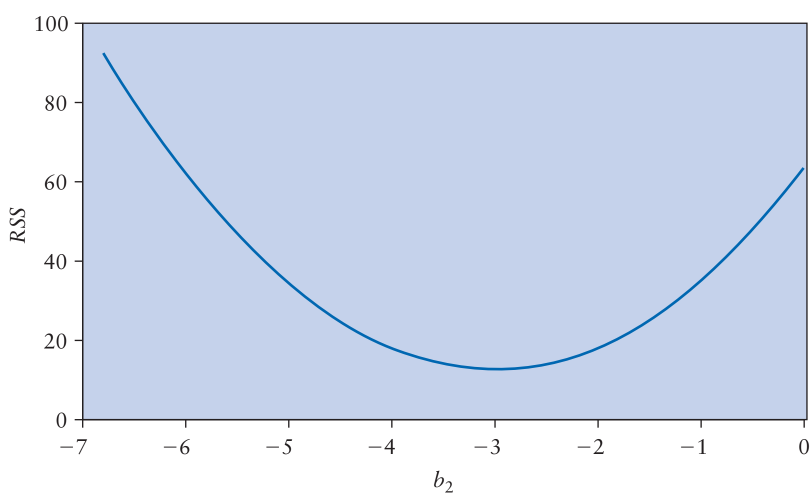

Figure 4.13 *RSS* as a function of b_2, conditional on $b_1 = 2$

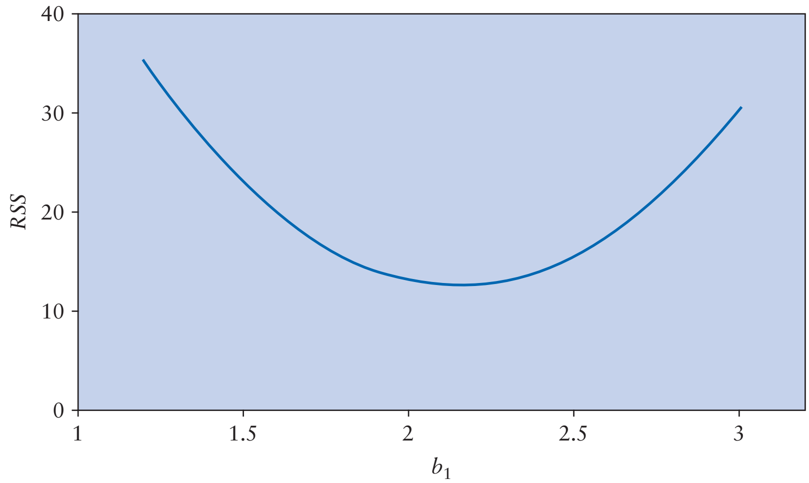

Figure 4.14 *RSS* as a function of b_1, conditional on $b_2 = -2.86$

Table 4.14

```
. nl (e = {beta1} + {beta2}/g)
(obs = 25)
Iteration 0: residual SS = 11.58161
Iteration 1: residual SS = 11.58161

    Source   |        SS          df          MS           Number of obs =       25
-------------+------------------------------------         R-squared     =   0.5311
     Model   |     13.1203672      1      13.1203672       Adj R-squared =   0.5108
  Residual   |     11.5816083     23      .503548186       Root MSE      = .7096113
-------------+------------------------------------         Res. dev.     = 51.71049
     Total   |     24.7019754     24      1.02924898
-----------------------------------------------------------------------------------
         e   |      Coef.     Std. Err.     t     P>|t|      [95% Conf. Interval]
-------------+---------------------------------------------------------------------
    /beta1   |     2.604753   .3748821    6.95    0.000      1.82925     3.380256
    /beta2   |    -4.050817    .793579   -5.10    0.000     -5.69246    -2.409174
-----------------------------------------------------------------------------------
Parameter beta1 taken as constant term in model & ANOVA table
```

In practice, the algorithms used for minimizing the residual sum of squares in a nonlinear model are mathematically far more sophisticated than the simple trial-and-error method described above. Nevertheless, until fairly recently a major problem with the fitting of nonlinear regressions was that it was very slow compared with linear regression, especially when there were several parameters to be estimated, and the high computing cost discouraged the use of nonlinear regression. This has changed as the speed and power of computers have increased. As a consequence, more interest is being taken in the technique and regression applications often incorporate user-friendly nonlinear regression features.

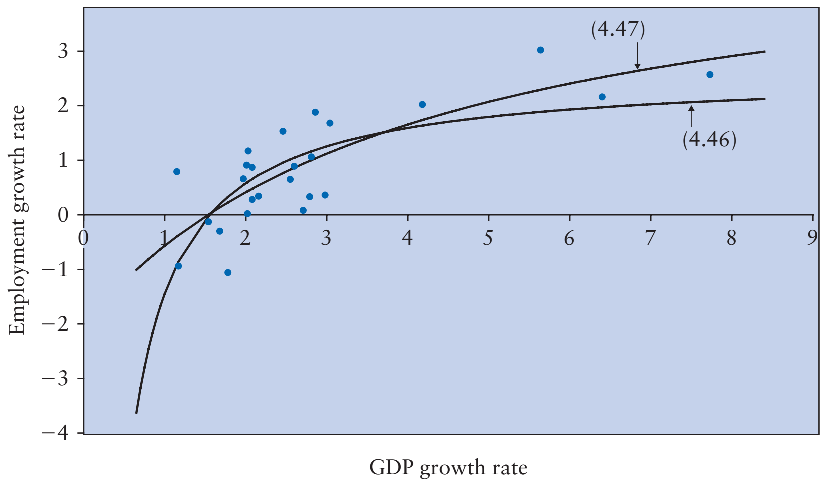

Figure 4.15 Alternative hyperbolic specifications (4.46) and (4.47)

Table 4.15

```
. nl (e = {beta1} + {beta2}/({beta3} + g))
(obs = 25)
Iteration 0: residual SS = 11.58161
Iteration 1: residual SS = 11.19238
 ------------------------------------------------------------------------------------
Iteration 15: residual SS = 9.01051

    Source   |        SS            df          MS          Number of obs =        25
-------------+-------------------------------------------   R-squared     =    0.6352
     Model   |     15.6914659        2      7.84573293      Adj R-squared =    0.6021
  Residual   |     9.01050957       22      .409568617      Root MSE      = .6399755
-------------+-------------------------------------------   Res. dev.     = 45.43482
     Total   |     24.7019754       24      1.02924898

----------------------------------------------------------------------------------
         e   |      Coef.      Std. Err.     t      P>|t|     [95% Conf. Interval]
-------------+--------------------------------------------------------------------
    /beta1   |     5.467548    2.826401    1.93    0.066     -.3940491    11.32914
    /beta2   |     -31.0764    41.78914   -0.74    0.465     -117.7418    55.58897
    /beta3   |     4.148589    4.870437    0.85    0.404      -5.95208    14.24926
----------------------------------------------------------------------------------
Parameter beta1 taken as constant term in model & ANOVA table
```

Table 4.14 shows such output for the present hyperbolic regression of *e* on *g*. It is, as usual, Stata output, but output from other regression applications will look similar. The Stata command for a nonlinear regression is ‘nl’. This is followed by the hypothesized mathematical relationship within parentheses. The parameters must be given names placed within braces. Here β_1 is {beta1} and β_2 is {beta2}. The output is effectively the same as the linear regression output in Table 4.2.

The hyperbolic function (4.46) imposes the constraint that the function plunges to minus infinity for positive g as g approaches zero. This feature can be relaxed by using the variation

$$e = \beta_1 + \frac{\beta_2}{\beta_3 + g} + u. \tag{4.47}$$

Unlike (4.46), this cannot be linearized by any kind of transformation. Here, nonlinear regression must be used. Table 4.15 gives the corresponding output, with most of the iteration messages deleted. Figure 4.15 compares the original and new hyperbolic functions. The fit is a considerable improvement, reflected in a higher R^2.

Key terms

- elasticity
- linear in parameters
- linear in variables
- logarithmic model
- logarithmic transformation
- loglinear model
- nonlinear regression algorithm
- semilogarithmic model

EXERCISE

4.16*

```
. nl (S = {beta1} + {beta2}/({beta3} + SIBLINGS)) if SIBLINGS>0
(obs = 529)
Iteration 0: residual SS = 2962.929
Iteration 1: residual SS = 2951.616
----------------------------------
Iteration 13: residual SS = 2926.201

  Source  |      SS          df         MS          Number of obs =      529
----------+-------------------------------------    R-squared     =   0.0659
   Model  |  206.566702       2     103.283351      Adj R-squared =   0.0624
Residual  |  2926.20078     526     5.56311936      Root MSE      = 2.358627
----------+-------------------------------------    Res. dev.     = 2406.077
   Total  |  3132.76749     528     5.93327175

------------------------------------------------------------------------------
       s  |     Coef.    Std. Err.     t     P>|t|     [95% Conf. Interval]
----------+-------------------------------------------------------------------
  /beta1  |   11.09973   1.363292    8.14    0.000     8.421565     13.7779
  /beta2  |   17.09479   18.78227    0.91    0.363    -19.80268    53.99227
  /beta3  |   3.794949    3.66492    1.04    0.301    -3.404729    10.99463
------------------------------------------------------------------------------
Parameter beta1 taken as constant term in model & ANOVA table
```

The output above uses *EAEF* Data Set 21 to fit the nonlinear model

$$S = \beta_1 + \frac{\beta_2}{\beta_3 + SIBLINGS} + u,$$

where S is the years of schooling of the respondent and *SIBLINGS* is the number of brothers and sisters. The specification is an extension of that for Exercise 4.1, with the addition of the parameter β_3. Provide an interpretation of the regression results and compare it with that for Exercise 4.1.