Chapter 10

Nonlinear Models

- Nonlinear models can be classified into two categories. In the first category are models that are nonlinear in the variables, but still linear in terms of the unknown parameters. This category includes models which are made linear in the parameters via a transformation.
- For example, the Cobb-Douglas production function that relates output (Y) to labor (L) and capital (K) can be written as

$$Y = \alpha L^{\beta} K^{\gamma}$$

Taking logarithms yields

$$\ln(Y) = \delta + \beta \ln(L) + \gamma \ln(K)$$

where $\delta = \ln(\alpha)$. This function is nonlinear in the variables Y, L, and K, but it is linear in the parameters δ, β and γ. Models of this kind can be estimated using the least-squares technique.

- The second category of nonlinear models contains models which are nonlinear in the parameters and which cannot be made linear in the parameters after a transformation. For estimating models in this category the familiar least squares technique is extended to an estimation procedure known as *nonlinear least squares*.

10.1 Polynomial and Interaction Variables

Models with polynomial and/or interaction variables are useful for describing relationships where the response to a variable changes depending on the value of that variable or the value of another variable. In contrast to the dummy variable examples in Chapter 9, we model relationships in which the slope of the regression model is *continuously* changing. We consider two such cases, interaction variables that are the product of a variable by itself, producing a polynomial term; and interaction variables that are the product of two different variables.

10.1.1 Polynomial Terms in a Regression Model

- In microeconomics you studied "cost" curves and "product" curves that describe a firm. Total cost and total product curves are mirror images of each other, taking the standard "cubic" shapes shown in Figure 10.1. Average and marginal cost curves, and their mirror images, average and marginal product curves, take quadratic shapes, usually represented as shown in Figure 10.2.

- The slopes of these relationships are not constant and cannot be represented by regression models that are "linear in the variables." However, these shapes are easily represented by polynomials, that are a special case of interaction variables in which variables are multiplied by themselves.
- For example, if we consider the average cost relationship in Figure 10.2*a*, a suitable regression model is:

$$AC = \beta_1 + \beta_2 Q + \beta_3 Q^2 + e \qquad (10.1.1)$$

This quadratic function can take the "U" shape we associate with average cost functions.

- For the total cost curve in Figure 10.1*a* a cubic polynomial is in order:

$$TC = \alpha_1 + \alpha_2 Q + \alpha_3 Q^2 + \alpha_4 Q^3 + e \qquad (10.1.2)$$

- These functional forms, which represent nonlinear shapes, *are still linear regression models, since the parameters enter in a linear way*. The variables Q^2 and Q^3 are explanatory variables that are treated no differently from any others. The parameters in Equations (10.1.1) and (10.1.2) can still be estimated by least squares.
- A difference in these models is in the interpretation of the parameters. The parameters of these models are not themselves slopes. The slope of the average cost curve (10.1.1) is

$$\frac{dE(AC)}{dQ} = \beta_2 + 2\beta_3 Q \qquad (10.1.3)$$

The slope of the average cost curve changes for every value of Q and depends on the parameters β_2 and β_3. For this U-shaped curve we expect $\beta_2 < 0$ and $\beta_3 > 0$. The slope of the total cost curve (10.1.2), which is the marginal cost, is

$$\frac{dE(TC)}{dQ} = \alpha_2 + 2\alpha_3 Q + 3\alpha_4 Q^2 \qquad (10.1.4)$$

The slope is a quadratic function of Q, involving the parameters α_2, α_3, and α_4. For a U-shaped marginal cost curve $\alpha_2 > 0$, $\alpha_3 < 0$, and $\alpha_4 > 0$.

- Using polynomial terms is an easy and flexible way to capture nonlinear relationships between variables. Their inclusion does not complicate least squares estimation. As we have shown, however, care must be taken when interpreting the parameters of models containing polynomial terms.

10.1.2 Interactions Between Two Continuous Variables

- When the product of two continuous variables is included in a regression model, the effect is to alter the relationship between each of them and the dependent variable. We will consider a "life-cycle" model to illustrate this idea.
- Suppose we wish to study the effect of income and age on an individual's expenditure on pizza. For this purpose we take a random sample of 40 individuals, age 18 and older, and record their annual expenditure on pizza (*PIZZA*), their income (*Y*) and age (*AGE*). The first 5 observations of these data are shown in Table 10.1.
- As an initial model consider

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 Y + e \tag{10.1.5}$$

The implications of this specification are:

1. $\dfrac{\partial E(PIZZA)}{\partial AGE} = \beta_2$: For a *given level of income*, the expected expenditure on pizza changes by the amount β_2 with an additional year of age. We expect the sign of β_2 to be negative. With the effects of income removed, we expect that as a person ages his/her pizza expenditure will fall.

2. $\dfrac{\partial E(PIZZA_i)}{\partial Y_i} = \beta_3$: For individuals of a *given age*, an increase in income of \$1 increases expected expenditures on pizza by β_3. Since pizza is probably a normal good, we expect the sign of β_3 to be positive. The parameter β_3 might be called the marginal propensity to spend on pizza.

- It seems unreasonable to expect that, *regardless* of the age of the individual, an increase in income by \$1 should lead to an increase in pizza expenditure by β_3 dollars. It would seem reasonable to assume that as a person grows older, their marginal propensity to spend on pizza declines. That is, as a person ages, less of each extra dollar is expected to be spent on pizza. This is a case in which *the effect of income depends on the age of the individual*. That is, the effect of one variable is modified by another.
- One way of accounting for such interactions is to include an interaction variable that is the product of the two variables involved. Since *AGE* and *Y* are the variables that interact, we will add the variable $(AGE \times Y)$ to the regression model. The result is

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 Y + \beta_4 (AGE \times Y) + e \qquad (10.1.6)$$

* When the product of two continuous variables is included in a model, the interpretation of the parameters requires care. The effects of *Y* and *AGE* are:

1. $\dfrac{\partial E(PIZZA)}{\partial AGE} = \beta_2 + \beta_4 Y$: The effect of *AGE* now depends on income. As a person ages his/her pizza expenditure is expected to fall, and, because β_4 is expected to be negative, the greater the income the greater will be the fall attributable to a change in age.

2. $\dfrac{\partial E(PIZZA)}{\partial Y} = \beta_3 + \beta_4 AGE$: The effect of a change in income on expected pizza expenditure, which is the marginal propensity to spend on pizza, now depends on *AGE*. If our logic concerning the effect of aging is correct, then β_4 should be negative. Then, as *AGE* increases, the value of the partial derivative declines.

- Estimates of models (10.1.5) and (10.1.6), with *t*-statistics in parentheses, are:

$$\begin{aligned} \widehat{PIZZA} = {} & 342.8848 - 7.5756AGE + 0.0024Y \\ & (4.740) \quad (-3.270) \quad (3.947) \end{aligned} \tag{R10.1}$$

and

$$\begin{aligned} \widehat{PIZZA} = {} & 161.4654 - 2.9774AGE + 0.0091Y - 0.00016(Y \times AGE) \\ & (1.338) \quad (-0.888) \quad (2.473) \quad (-1.847) \end{aligned} \tag{R10.2}$$

- In (R10.1) the signs of the estimated parameters are as we anticipated. Both *AGE* and income (*Y*) have significant coefficients, based on their *t*-statistics. In (R10.2) the product (*AGE* × *Y*) enters the equation. Its estimated coefficient is negative and significant at the α = .05 level using a one-tailed test. The signs of other coefficients remain the same, but *AGE*, by itself, no longer appears to be a significant explanatory factor. This suggests that *AGE* affects pizza expenditure through its interaction with income—that is, it affects the marginal propensity to spend on pizza.

- Using the estimates in (R10.2) let us estimate the marginal effect of age upon pizza expenditure for two individuals; one with \$25,000 income and one with \$90,000 income.

$$\frac{\partial E(P\hat{I}ZZA)}{\partial AGE} = b_2 + b_4 Y = -2.9774 - 0.00016Y$$
$$= \begin{cases} -6.9774 & \text{for } Y = \$25,000 \\ -17.3774 & \text{for } Y = \$90,000 \end{cases} \tag{R10.3}$$

That is, we expect that an individual with $25,000 income will reduce expenditure on pizza by $6.98 per year, while the individual with $90,000 income will reduce pizza expenditures by $17.38 per year, all other factors held constant.

10.2 A Simple Nonlinear-in-the-Parameters Model

We turn now to models that are nonlinear in the parameters and which need to be estimated by a technique called nonlinear least squares. There are a variety of models that fit into this framework, because of the functional form of the relationship being modeled, or because of the statistical properties of the variables.

- To explain the nonlinear least estimation technique, we consider the following artificial example

$$y_t = \beta x_{t1} + \beta^2 x_{t2} + e_t \tag{10.2.1}$$

where y_t is a dependent variable, x_{t1} and x_{t2} are explanatory variables, β is an unknown parameter that we wish to estimate, and the e_t are uncorrelated random errors with mean zero and variance σ^2. This example differs from the conventional linear model because the coefficient of x_{t2} is equal to the square of the coefficient x_{t1}.

- When we had a simple linear regression equation with two unknown parameters β_1 and β_2 we set up a sum of squared errors function. In the context of Equation (10.2.1),

$$S(\beta) = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (y_t - \beta x_{t1} - \beta^2 x_{t2})^2 \tag{10.2.2}$$

- When we have a nonlinear function like Equation (10.2.1), we *cannot* derive an algebraic expression for the parameter β that minimizes Equation (10.2.2). However, for a given set of data, we can ask the computer to look for the parameter value that takes us to the bottom of the bowl. Many software algorithms can be used to find *numerically* the value that minimizes S(β). This value is called a **nonlinear least squares estimate.**

- It is also impossible to get algebraic expressions for standard errors, but it is possible for the computer to calculate a numerical standard error. Estimates and standard errors computed in this way have good properties in large samples.
- As an example, consider the data on y_t, x_{t1}, and x_{t2} in Table 10.2. The sum of squared errors function in Equation (10.2.2) is graphed in Figure 10.2. Because we have only one unknown parameter, we have a two-dimensional curve, not a "bowl." It is clear that the minimizing value for β lies between 1.0 and 1.5.
- Using nonlinear least squares software, we find that the nonlinear least squares estimate and its standard error are

$$b = 1.1612 \qquad \text{se}(b) = 0.129 \qquad \text{(R10.4)}$$

- Be warned that different software can yield slightly different approximate standard errors. However, the nonlinear least squares estimate should be the same for all packages.

10.3 A Logistic Growth Curve

- A model that is popular for modelling the diffusion of technological change is the logistic growth curve

$$y_t = \frac{\alpha}{1 + \exp(-\beta - \delta t)} + e_t \qquad (10.3.1)$$

- In the above equation y_t is the adoption proportion of a new technology. In our example y_t is the share of total U.S. crude steel production that is produced by electric arc furnace technology.
- There is only one explanatory variable on the right hand side, namely, time, $t = 1, 2,\ldots,T$. Thus, the logistic growth model is designed to capture the rate of adoption of technological change, or, in some examples, the rate of growth of market share.

- An example of a logistic curve is depicted in Figure 10.4. The rate of growth increases at first, to a point of inflection which occurs at $t = -\beta/\delta = 20$. Then, the rate of growth declines, leveling off to a saturation proportion given by $\alpha = 0.8$.

- Since $y_0 = \alpha/(1 + \exp(-\beta))$, the parameter β determines how far the share is below saturation level at time zero. The parameter δ controls the speed at which the point of inflection, and the saturation level, are reached. The curve is such that the share at the point of inflection is $\alpha/2 = 0.4$, half the saturation level.

- The e_t are assumed to be uncorrelated random errors with zero mean and variance σ^2. Because the parameters in Equation (10.3.1) enter the equation in a nonlinear way, it is estimated using nonlinear least squares.

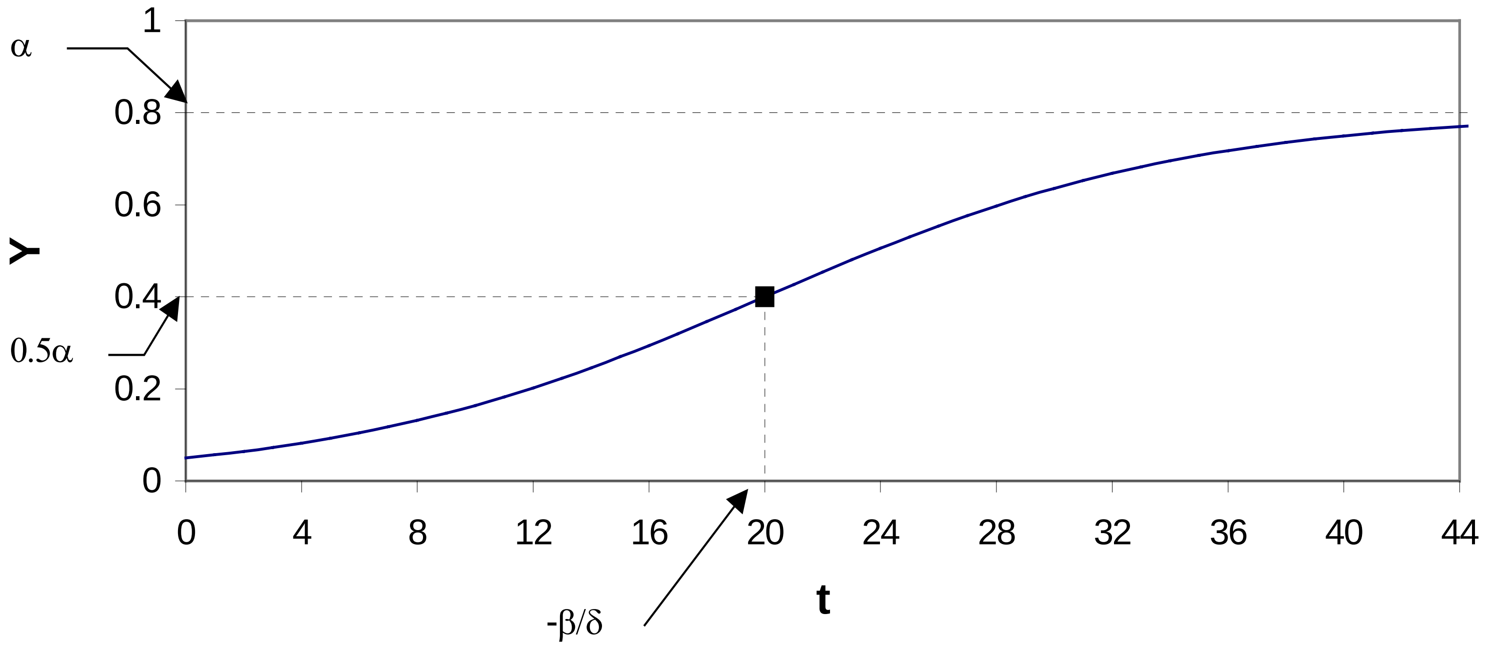

Figure 10.4 **Logistic Growth Curve**

- To illustrate estimation of Equation (10.3.1) we use data on the electric arc furnace (EAF) share of steel production in the U.S. These data appear in Table 10.3.
- Using nonlinear least squares to estimate the logistic growth curve yields the results in Table 10.4. We find that the estimated saturation share of the EAF technology is $\hat{\alpha} = 0.46$. The point of inflection, where the rate of adoption changes from increasing to decreasing, is estimated as

$$-\frac{\hat{\beta}}{\hat{\delta}} = \frac{0.911}{0.117} = 7.8 \qquad \text{(R10.5)}$$

which is approximately the year 1977.

- In the upper part of Table 10.4 is the phrase "convergence achieved after 8 iterations." This means that the numerical procedure used to minimize the sum of squared errors took 8 steps to find the minimizing least squares estimates. If you run a nonlinear least squares problem and your software reports that convergence has not occurred, you should not use the "estimates" from that run.

- Suppose that you wanted to test the hypothesis that the point of inflection actually occurred in 1980. The corresponding null and alternative hypotheses can be written as H_0: $-\beta/\delta = 11$ and H_1: $-\beta/\delta \neq 11$, respectively.

- The null hypothesis is different from any that you have encountered so far because it is nonlinear in the parameters β and δ. Despite this nonlinearity, the test can be carried out using most modern software. The outcome of this test appears in the last two rows of Table 10.4 under the heading "Wald test." From the very small *p*-values associated with both the F and the χ^2-statistics, we reject H_0 and conclude that the point of inflection does not occur at 1980.

Table 10.4 **Estimated Growth Curve for EAF Share of Steel Production.**

Dependent Variable: Y
Method: Least Squares
Date: 11/20/99 Time: 15:19
Sample: 1970 1997
Included observations: 28
Convergence achieved after 8 iterations
Y=C(1)/(1+EXP(-C(2)-C(3)*T))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.462303	0.018174	25.43765	0.0000
C(2)	-0.911013	0.058147	-15.66745	0.0000
C(3)	0.116835	0.010960	10.65979	0.0000

Wald Test:
Null Hypothesis: -C(2)/C(3)=11

F-statistic	16.65686	Probability	0.000402
Chi-square	16.65686	Probability	0.000045

10.4 Poisson Regression

- To help decide the annual budget allocations for recreational areas, the State Government collects information on the demand for recreation. It took a random sample of 250 households from households who live within a 120 mile radius of Lake Keepit. Households were asked a number of questions, including how many times they visited Lake Keepit during the last year.

- The frequency of visits appears in Table 10.5. Note the special nature of the data in this table. There is a large number of households who did not visit the Lake at all, and also large numbers for 1 visit, 2 visits and 3 visits. There are fewer households who made a greater number of trips, such as 6 or 7.

Table 10.5 Frequency of Visits to Keepit Dam

Number of visits	0	1	2	3	4	5	6	7	8	9	10	13
Frequency	61	55	41	31	23	19	8	7	2	1	1	1

- Data of this kind are called *count data*. The possible values that can occur are the countable integers 0, 1, 2, … . Count data can be viewed as observations on a *discrete random variable*. A distribution suitable for count data is the *Poisson distribution* rather than the normal distribution. Its probability density function is given by

$$f(y) = \frac{\mu^{y} \exp(-\mu)}{y!} \qquad (10.4.1)$$

- In the context of our example, y is the number of times a household visits Lake Keepit per year and μ is the average or mean number of visits per year, for all households. Recall that $y! = y \times (y - 1) \times (y - 2) \times \ldots \times 2 \times 1$.

- In *Poisson regression*, we improve on Equation (10.4.1) by recognizing that the mean μ is likely to depend on various household characteristics. Households who live close to the lake are likely to visit more often than more-distant households. If recreation is a normal good, the demand for recreation will increase with income. Larger household (more family members) are likely to make more frequent visits to the lake. To accommodate these differences, we write μ_i, the mean for the ith household as

$$\mu_i = \exp(\beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}) \tag{10.4.2}$$

where the β_j's are unknown parameters and

x_{i2} = distance of the *i*-th household from the Lake in miles,

x_{i3} = household income in tens of thousands of dollars, and

x_{i4} = number of household members.

Writing μ_i as an exponential function of x_2, x_3, and x_4, rather than a simple linear function, ensures μ_i will be positive.

- Recall that, in the simple *linear* regression model, we can write

$$y_i = \mu_i + e_i = \beta_1 + \beta_2 x_i + e_i \qquad (10.4.3)$$

The mean of y_i is $\mu_i = E(y_i) = \beta_1 + \beta_2 x_i$. Thus, μ_i can be written as a function of the explanatory variable x_i. The error term e_i is defined as $y_i - \mu_i$, and, consequently, has a zero mean.

- We can proceed in the same way with our Poisson regression model. We define the zero-mean error term $e_i = y_i - \mu_i$, or $y_i = \mu_i + e_i$, from which we can write

$$y_i = \exp(\beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}) + e_i \tag{10.4.4}$$

Equation (10.4.4) can be estimated via nonlinear least squares since it is nonlinear in the parameters. Estimating the equation tells us how the demand for recreation at Lake Keepit depends on distance traveled, income, and number of household numbers. It also gives us a model for predicting the number of visitors to Lake Keepit.

- The nonlinear least squares estimates of Equation (10.4.4) appear in Table 10.6. Because of the nonlinear nature of the function, we must be careful how we interpret the magnitudes of the coefficients.

- However, examining their signs, we can say the greater the distance from Lake Keepit, the less will be the expected number of visits. Increasing income, or the size of the household, increases the frequency of visits. The income coefficient is not significantly different from zero, but those for distance and household members are.

Table 10.6 **Estimated Model for Visits to Lake Keepit**

Dependent Variable: VISITS
Method: Least Squares
Date: 11/20/99 Time: 09:15
Sample: 1 250
Included observations: 250
Convergence achieved after 7 iterations
VISITS=EXP(C(1)+C(2)*DIST+C(3)*INC+C(4)*MEMB)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.390670	0.176244	7.890594	0.0000
C(2)	-0.020865	0.001749	-11.93031	0.0000
C(3)	0.022814	0.015833	1.440935	0.1509
C(4)	0.133560	0.030310	4.406527	0.0000

- The estimated model can also be used to compute probabilities relating to a household with particular characteristics. For example, what is the probability that a household located 50 miles from the Lake, with income of \$60,000, and 3 family members, visits the park less than 3 times per year? First we compute an estimate of the mean for this household

$$\begin{aligned}\hat{\mu} &= \exp(1.39067 - 0.020865\times 50 + 0.022814\times 6 + 0.13356\times 3) \\ &= 2.423\end{aligned} \tag{R10.6}$$

Then, using the Poisson distribution, we have

$$
\begin{aligned}
P(y<3) &= P(y=0)+P(y=1)+P(y=2) \\
&= \frac{(2.423)^0 \exp(-2.423)}{0!} + \frac{(2.423)^1 \exp(-2.423)}{1!} \\
&\quad + \frac{(2.423)^2 \exp(-2.423)}{2!} \\
&= 0.0887 + 0.2148 + 0.2602 \\
&= 0.564
\end{aligned}
\tag{R10.7}
$$

Other probabilities can be computed in a similar way.

Exercise

10.1	10.3	10.4	10.5	10.6
10.8	10.9			