Lecture 10: Nonlinear Regression Functions

Zheng Tian

Outline

1. Introduction
2. A General Strategy For Modeling Nonlinear Regression Functions
3. Nonlinear functions of a single independent variable
4. Interactions between independent variables
5. Regression Functions That Are Nonlinear in the Parameters

Overview

Linear population regression function

$E(Y_i \mid \mathbf{X}_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})'$.

Nonlinear population regression function

$E(Y_i \mid \mathbf{X}_i) = f(X_{i1}, X_{i2}, \ldots, X_{ik}; \beta_1, \beta_2, \ldots, \beta_m)$, where $f(\cdot)$ is a nonlinear function.

Study questions

- Why do we need to use nonlinear regression models?
- What types of nonlinear regression models can we estimate by OLS?
- How can we interpret the coefficients in nonlinear regression models?

Test Scores and district income

- Test scores can be determined by average district income
- We estimate a simple linear regression model

$$TestScore = \beta_0 + \beta_1 Income + u$$

- What's the problem with the simple linear regression model?

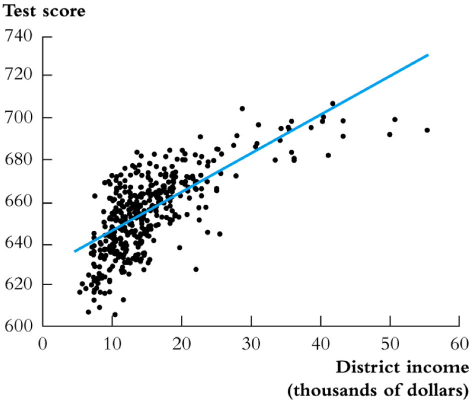


Figure: Scatterplot of test score vs district income and a linear regression line

Why does a simple linear regression model not fit the data well?

- Data points are below the OLS line when income is very low (under \$10,000) or very high (over \$40,000), and are above the line when income is between \$15,000 and \$30,000.
- The scatterplot may imply a curvature in the relationship between test scores and income.

  That is, a unit increase in income may have larger effect on test scores when income is very low than when income is very high.
- The linear regression line cannot capture the curvature because the effect of district income on test scores is constant over all the range of income since

$$\Delta TestScore / \Delta Income = \beta_1$$

  where β_1 is constant.

Estimate a quadratic regression model

$$TestScore = \beta_0 + \beta_1 Income + \beta_2 Income^2 + u \quad (1)$$

- This model is nonlinear, specifically quadratic, with respect to *Income* since we include the squared income.
- The population regression function is

$$E(TestScore|Income) = \beta_0 + \beta_1 Income + \beta_2 Income^2$$

- It is linear with respect to β. So we can still use the OLS estimation and carry out hypothesis testing as we do with a linear regression model.

Estimate a quadratic regression model (cont'd)

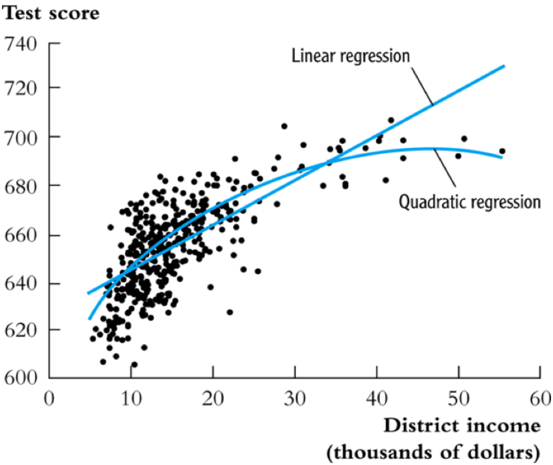


Figure: Scatterplot of test score vs district income and a quadratic regression line

A general formula for a nonlinear population regression function

A general nonlinear regression model is

$$Y_i = f(X_{i1}, X_{i2}, \ldots, X_{ik}; \beta_1, \beta_2, \ldots, \beta_m) + u_i \tag{2}$$

- The population nonlinear regression function:

$$E(Y_i|X_{i1}, \ldots, X_{ik}) = f(X_{i1}, X_{2i}, \ldots, X_{ik}; \beta_1, \beta_2, \ldots, \beta_m)$$

- The number of regressors and the number of parameters are not necessarily equal in the nonlinear regression model.
- In vector notation

$$Y_i = f(\mathbf{X}_i; \boldsymbol{\beta}) + u_i \tag{3}$$

- We focus on the nonlinear regression models such that $f(\cdot)$ is nonlinear with $\mathbf{X}_i$ but linear with $\boldsymbol{\beta}$.

The effect on Y of a change in a regressor

For any general nonlinear regression function

The effect on Y of a change in one regressor, say X_1, holding other things constant, can be computed as

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \ldots, X_k; \boldsymbol{\beta}) - f(X_1, X_2, \ldots, X_k; \boldsymbol{\beta}) \tag{4}$$

For continuous and differentiable nonlinear functions

When X_1 and Y are continuous variables and $f(\cdot)$ is differentiable, the marginal effect of X_1 is the partial derivative of f with respect to X_1, that is, holding other things constant

$$\mathrm{d}Y = \frac{\partial f(X_1, \ldots, X_k; \boldsymbol{\beta})}{\partial X_i} \mathrm{d}X_i$$

because $\mathrm{d}X_j = 0$ for $j \neq i$

Application to test scores and income

Estimation

$$\widehat{TestScore} = \underset{(2.9)}{607.3} + \underset{(0.27)}{3.85}\, Income - \underset{(0.0048)}{0.0423}\, Income^2,\ \bar{R}^2 = 0.554 \tag{5}$$

Hypothesis test

Test $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$.

$$t = \frac{-0.0423}{0.0048} = -8.81 > -1.96$$

We reject the null at the 1%, 5% and 10% significance levels, and therefore, confirm the quadratic relationship between test scores and income.

The effect of change in income on test scores

A change in income from $10 thousand to $20 thousand

$$\begin{aligned}\Delta \hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 \times 11 + \hat{\beta}_2 \times 11^2 - (\hat{\beta}_0 + \hat{\beta}_1 \times 10 + \hat{\beta}_2 \times 10^2) \\ &= \hat{\beta}_1(11 - 10) + \hat{\beta}_2(11^2 - 10^2) \\ &= 3.85 - 0.0423 \times 21 = 2.96\end{aligned}$$

A change in income from $40 thousand to $41 thousand

$$\begin{aligned}\Delta \hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 \times 41 + \hat{\beta}_2 \times 41^2 - (\hat{\beta}_0 + \hat{\beta}_1 \times 40 + \hat{\beta}_2 \times 40^2) \\ &= \hat{\beta}_1(41 - 40) + \hat{\beta}_2(41^2 - 40^2) \\ &= 3.85 - 0.0423 \times 81 = 0.42\end{aligned}$$

A general approach to modeling nonlinearities using multiple regression

1. Identify a possible nonlinear relationship.
    - Economic theory
    - Scatterplots
    - Your judgment and experts' opinions
2. Specify a nonlinear function and estimate its parameters by OLS.
    - The OLS estimation and inference techniques can be used as usual when the regression function is linear with respect to β.
3. Determine whether the nonlinear model can improve a linear model
    - Use t- and/or F-statistics to test the null hypothesis that the population regression function is linear against the alternative that it is nonlinear.
4. Plot the estimated nonlinear regression function.
5. Compute the effect on Y of a change in X and interpret the results.

Polynomials

A polynomial regression model of degree r

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_r X_i^r + u_i \tag{6}$$

- $r = 2$: a quadratic regression model
- $r = 3$: a cubic regression model
- Use the OLS method to estimate $\beta_1, \beta_2, \ldots, \beta_r$.

Testing the null hypothesis that the population regression function is linear

$$H_0: \beta_2 = 0, \beta_3 = 0, ..., \beta_r = 0 \text{ vs. } H_1: \text{ at least one } \beta_j \neq 0, j = 2, \ldots, r$$

Use F statistic to test this joint hypothesis. The number of restriction is $q = r - 1$.

What is $\Delta Y / \Delta X$ in a polynomial regression model?

- Consider a cubic model and continuous X and Y

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + u$$

- Then, we can calculate

$$\frac{\mathrm{d}Y}{\mathrm{d}X} = \beta_1 + 2\beta_2 X + 3\beta_3 X^2$$

- The effect of a unit change in X on Y depends on the value of X at evaluation.

Which degree of polynomial should I use?

- Balance a trade-off between flexibility and statistical precision.
    - Flexibility. Relate Y to X in more complicated way than simple linear regression.
    - Statistical precision. $X, X^2, X^3, \ldots$ are correlated so that there is the problem of imperfect multicollinearity.
- Follow a sequential hypothesis testing procedure
    1. Pick a maximum value of r and estimate the polynomial regression for that r.
    2. Follow a "deletion" rule based on t-statistic or F-statistic.

Application to district income and test scores

We estimate a cubic regression model relating test scores to district income as follows

$$\widehat{TestScore} = \underset{(5.1)}{600.1} + \underset{(0.71)}{5.02}\ Income - \underset{(0.029)}{0.096}\ Income^2 + \underset{(0.00035)}{0.00069}\ Income^3, \bar{R}^2 = 0.555$$

Test whether it is a cubic model

The t-statistic for $H_0 : \beta_3 = 0$ is 1.97 $\Rightarrow$ Fail to reject

Test whether it is a nonlinear model

The F-statistic for $H_0 : \beta_2 = \beta_3 = 0$ is 37.7, p-value < 0.01

Interpretation of coefficients

Use the general formula of interpreting the effect of ΔX on Y.

A natural logarithmic function $y = \ln(x)$

- Properties of $\ln(x)$

$$\ln(1/x) = -\ln(x), \ln(ax) = \ln(a) + \ln(x)$$
$$\ln(x/a) = \ln(x) - \ln(a), \text{ and } \ln(x^a) = a\ln(x)$$

- The derivative of $\ln(x)$ is

$$\frac{\mathrm{d}\ln(x)}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\ln(x + \Delta x) - \ln(x)}{\Delta x} = \frac{1}{x}.$$

It follows that $\mathrm{d}\ln(x) = \mathrm{d}x/x$, representing the percentage change in x.

The percentage-change form using $\ln(x)$

- The change in $\ln(X)$ represents the percentage change in X

$$\ln(x+\Delta x) - \ln(x) \approx \frac{\Delta x}{x} \text{ when } \Delta x \text{ is small.}$$

- The Taylor expansion of $\ln(x+\Delta x)$ at x, which is

$$\begin{aligned}\ln(x+\Delta x) &= \ln(x) + \frac{\mathrm{d}\ln(x)}{\mathrm{d}x}(x+\Delta x - x) + \frac{1}{2!}\frac{\mathrm{d}^2\ln(x)}{\mathrm{d}x^2}(x+\Delta x - x)^2 + \cdots \\ &= \ln(x) + \frac{\Delta x}{x} - \frac{\Delta x^2}{2x^2} + \cdots\end{aligned}$$

When Δx is very small, we can omit the terms with $\Delta x^2, \Delta x^3$, etc. Thus, we have $\ln(x+\Delta x) - \ln(x) \approx \frac{\Delta x}{x}$ when Δx is small.

The three logarithmic regression models

There are three types of logarithmic regression models:

- Linear-log model
- Log-linear model
- Log-log model

Differences in logarithmic transformation of X and/or Y lead to differences in interpretation of the coefficient.

Case I: linear-log model

- Model form. X is in logarithms, Y is not.

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i, i = 1, \ldots, n \tag{7}$$

- Interpretation. a 1% change in X is associated with a change in Y of $0.01\beta_1$

$$\Delta Y = \beta_1 \ln(X + \Delta X) - \beta_1 \ln(X) \approx \beta_1 \frac{\Delta X}{X}$$

- Example. The estimated model is

$$\widehat{TestScore} = 557.8 + 36.42 \ln(Income)$$

  - 1% increase in average district income results in an increase in test scores by $0.01 \times 36.42 = 0.36$ point.

Case II: log-linear model

- Model form. Y is in logarithms, X is not.

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i \tag{8}$$

- Interpretation. A one-unit change in X is associated with a $100 \times \beta_1\%$ change in Y because

$$\frac{\Delta Y}{Y} \approx \ln(Y + \Delta Y) - \ln(Y) = \beta_1 \Delta X$$

- Example.

$$\ln(\widehat{Earnings}) = 2.805 + 0.0087 Age$$

  - Earnings are predicted to increase by 0.87% for each additional year of age.

Case III: log-log model

- Model form. Both X and Y are in logarithms.

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i \tag{9}$$

- Interpretation: elasticity. 1% change in X is associated with a β_1% change in Y because

$$\frac{\Delta Y}{Y} \approx \ln(Y + \Delta Y) - \ln(Y) = \beta_1(\ln(X + \Delta X) - \ln(X)) \approx \beta_1 \frac{\Delta X}{X}$$

  - β_1 is the elasticity of Y with respect to X, that is

$$\beta_1 = \frac{100 \times (\Delta Y/Y)}{100 \times (\Delta X/X)} = \frac{\text{percentage change in } Y}{\text{percentage change in } X}$$

  - With the derivative, $\beta_1 = \mathrm{d}\ln(Y)/\mathrm{d}\ln(X) = (\mathrm{d}Y/Y)/(\mathrm{d}X/X)$.

- Example. The log-log model of the test score application is estimated as

$$\ln(\widehat{TestScore}) = 6.336 + 0.0544 \ln(Income)$$

This implies that a 1% increase in income corresponds to a 0.0544% increase in test scores.

The log-linear and log-log regression functions

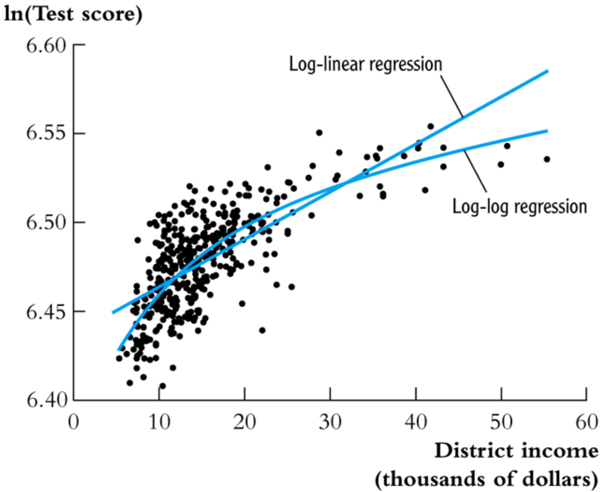


Figure: The log-linear and log-log regression functions

Summary

Regression specification	Interpretation of β_1
$Y = \beta_0 + \beta_1 \ln(X) + u$	A 1% change in X is associated with a change in Y of $0.01\beta_1$
$\ln(Y) = \beta_0 + \beta_1 X + u$	A change in X by one unit is associated with a $100\beta_1\%$ change in Y
$\ln(Y) = \beta_0 + \beta_1 \ln(X) + u$	A 1% change in X is associated with a $\beta_1\%$ change in Y, so β_1 is the elasticity of Y with respect to X

Interactions between independent variables

- Interaction between two binary variables
- Interaction between a continuous and a binary variable
- Interaction between two continuous variables

The regression model with interaction between two binary variables

Two binary variables

- $D_{1i} = 1$ if the i^{th} person has a college degree, and 0 otherwise.
- $D_{2i} = 1$ if the i^{th} person is female, and 0 otherwise.

A regression with an interaction term of two binary variables

Consider a regresion model concerning the effects of education and gender on earnings. The population regression function is

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i \tag{10}$$

- The dependent variable: Y_i, where $Y_i = Earnings_i$
- $D_{1i} \times D_{2i}$ is the interaction term.

The method of interpreting coefficients in regressions with interacted binary variables

We can follow a general rule for interpreting coefficients in Equation (10):

- First compute the expected values of Y for each possible case described by the set of binary variables.
- Next compare these expected values. Each coefficient can then be expressed either as an expected value or as the difference between two or more expected values.

Compute the expected values of Y for each possible combinations of D_1 and D_2

Case 1 $E(Y_i|D_{1i}=0, D_{2i}=0)=\beta_0$: the average income of male non-college graduates.

Case 2 $E(Y_i|D_{1i}=1, D_{2i}=0)=\beta_0+\beta_1$: the average income male college graduates.

Case 3 $E(Y_i|D_{1i}=0, D_{2i}=1)=\beta_0+\beta_2$: the average income of female non-college graduates.

Case 4 $E(Y_i|D_{1i}=1, D_{2i}=1)=\beta_0+\beta_1+\beta_2+\beta_3$: the average income of female college graduates.

Compute the difference between a pair of cases

Case 1 vs. Case 2 $E(Y_i|D_{1i}=1, D_{2i}=0) - E(Y_i|D_{1i}=0, D_{2i}=0) = \beta_1$: the average income difference between college graduates and non-college graduates among male workers.

Case 1 vs. Case 3 $E(Y_i|D_{1i}=0, D_{2i}=1) - E(Y_i|D_{1i}=0, D_{2i}=0) = \beta_2$: the average income difference between female and male workers who are not college graduates.

Case 1 vs. Case 4
$E(Y_i|D_{1i}=1, D_{2i}=1) - E(Y_i|D_{1i}=0, D_{2i}=0) = \beta_1 + \beta_2 + \beta_3$: the average income difference between female college graduates and male non-college graduates.

Compute the difference between a pair of cases (cont'd)

Case 2 vs. Case 3 $E(Y_i|D_{1i}=0, D_{2i}=1) - E(Y_i|D_{1i}=1, D_{2i}=0) = \beta_2 - \beta_1$.
Thus, the average income difference between female non-college graduates and male college graduates is $\beta_2 - \beta_1$.

Case 2 vs. Case 4 $E(Y_i|D_{1i}=1, D_{2i}=1) - E(Y_i|D_{1i}=1, D_{2i}=0) = \beta_2 + \beta_3$.
Thus, the average income difference between female college graduates and male college graduates is $\beta_2 + \beta_3$.

Case 3 vs. Case 4 $E(Y_i|D_{1i}=1, D_{2i}=1) - E(Y_i|D_{1i}=0, D_{2i}=1) = \beta_1 + \beta_3$.
Thus, the average income difference between female college graduates and female non-college graduates is $\beta_1 + \beta_3$.

Hypothesis testing

We can use t-statistic or F-statistic to test whether the differences between different cases are statistically significant.

The null hypothesis: $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$.

- What is this test for?
- What test statistic can we use?

The hypothesis is $H_0 : \beta_1 + \beta_3 = 0$ vs. $H_1 : \beta_1 + \beta_3 \neq 0$.

- What is this test for?
- What test statistic can we use?

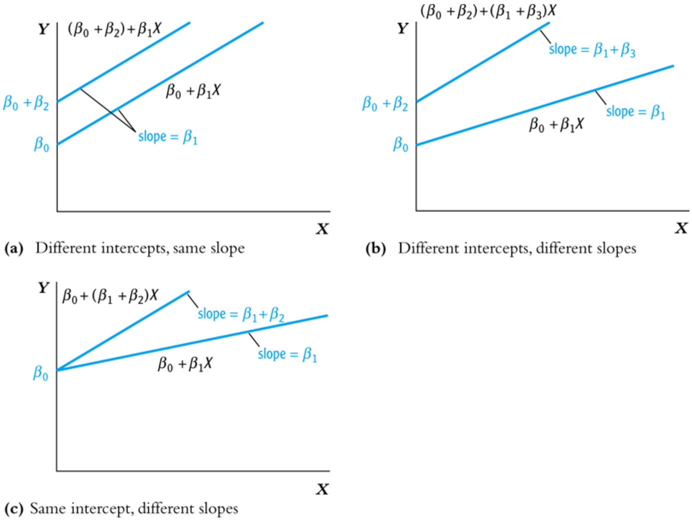
Interactions between a continuous and a binary variable

Consider the population regression of earnings (Y_i) against

- one continuous variable, individual's years of work experience (X_i), and
- one binary variable, whether the worker has a college degree (D_i, where $D_i = 1$ if the i$^{\text{th}}$ person is a college graduate).

As shown in the next figure, the population regression line relating Y and X can depend on D in three different ways.

Interactions between a continuous and a binary variable: graphic representation



Different intercept, same slope: (a) in Figure 4

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i \tag{11}$$

- From Equation (11), we have the population regression functions as
  - $E(Y_i|D_i = 1) = (\beta_0 + \beta_2) + \beta_1 X_i$
  - $E(Y_i|D_i = 0) = \beta_0 + \beta_1 X_i$.

  Thus, $E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \beta_2$.
- The average initial salary of college graduates is higher than non-college graduates by β_2, and this gap persists at the same magnitude regardless of how many years a worker has been working.

Different intercepts and different slopes: (b) in Figure 4

Equation (11):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i \tag{12}$$

- The population regression functions for the two cases are
  - $E(Y_i|D_i = 1) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_i$
  - $E(Y_i|D_i = 0) = \beta_0 + \beta_1 X_i$.

  Thus, β_2 is the difference in intercepts and β_3 is the difference in slopes.
- The average initial salary of college graduates is higher than non-college graduates by β_2, and this gap will widen (or narrow) depending on the effect of the years of work experience on earnings.

Different intercepts and same intercept: (c) in Figure 4

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i \times D_i) + u_i \tag{13}$$

- The population regression functions for the two cases are
  - $E(Y_i|D_i = 1) = \beta_0 + (\beta_1 + \beta_2)X_i$
  - $E(Y_i|D_i = 0) = \beta_0 + \beta_1 X_i$.

  Thus, there is only a difference in the slope but not in the intercept.
- Although college graduates have the same starting salary as those withou colledge degree, the raise in salary and promotion of the former will be faster than the latter.

Interactions between two continuous variables

Now we consider the regression of earnings against two continuous variables, one for the years of work experience (X_1) and another for the years of schooling (X_2).

The interaction model is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i \tag{14}$$

- The effect of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$$

- Similarly, the effect of a change in X_2, holding X_1 constant, is

$$\frac{\Delta Y}{\Delta X_2} = \beta_1 + \beta_3 X_1$$

Nonlinear regression models and nonlinear least squares estimator

All the regression models that we have discussed in this lecture are nonlinear in the regressors but linear in parameters so that we can still treat them as linear regression models and estimate using the OLS.

However, there exist regression models that are nonlinear in parameters. For these models, we can either transform them to the "linear" type of models or estimate using the nonlinear least squares (NLS) estimators.

Transform a nonlinear model to a linear one

Suppose we have a nonlinear regression model as follows

$$Y_i = \alpha X_{1i}^{\beta_1} X_{2i}^{\beta_2} \cdots X_{ki}^{\beta_k} e^{u_i} \tag{15}$$

Taking the natural logarithmic function on both sides of the equation

$$\ln(Y_i) = \ln(\alpha) + \beta_1 \ln(X_{1i}) + \beta_2 \ln(X_{2i}) + \cdots + \beta_k \ln(X_{ki}) + u_i \tag{16}$$

- Equation (15) becomes a log-log regression model, which is linear in all parameters and can be estimated using the OLS. Let $\beta_0 = \ln(\alpha)$ and $\alpha = e^{\beta_0}$.
- β_i for $i = 1, 2, \ldots, k$ are the elasticities of Y with respect to X_i.

A nonlinear model: logistic function

- A dependent variable can only take values between 0 and 1.
- The logistic regression model with k regressors is

$$Y_i = \frac{1}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \cdots \beta_k X_{ki})} + u_i \tag{17}$$

  - For small values of X, the value of the function is nearly 0 and the shape is flat.
  - For large values of X, the function approaches 1 and the slope is flat again.

A nonlinear model: negative exponential growth function

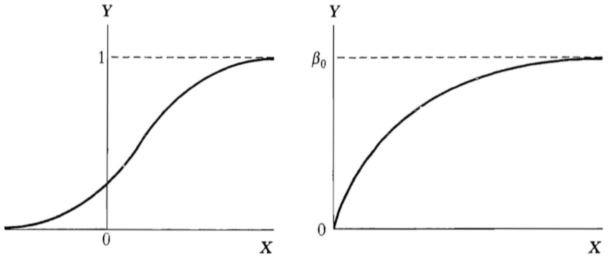
- The effect of X on Y must be positive and the effect is bounded by a upper bound.
- Use the negative-exponential growth function to set up a regression model as follows

$$Y_i = \beta_0[1 - \exp(-\beta_1(X_i - \beta_2))] + u_i \tag{18}$$

  - The slope is positive for all values of X.
  - The slope is greatest at low values of X and decreases as X increases.
  - There is an upper bound, that is, a limit of Y as X goes to infinity, β_0.

Nonlinear models that cannot be linearized \ Logistic and negative exponential growth curves

FIGURE 8.12 Two Functions That Are Nonlinear in Their Parameters



(a) A logistic curve

(b) A negative exponential growth curve

Part (a) plots the logistic function of Equation (8.38), which has predicted values that lie between 0 and 1. Part (b) plots the negative exponential growth function of Equation (8.39), which has a slope that is always positive and decreases as X increases, and an asymptote at β_0 as X tends to infinity.

Figure: The logistic and negative exponential growth functions

The nonlinear least squares estimators

For a nonlinear regression function

$$Y_i = f(X_1, \dots, X_k; \beta_1, \dots, \beta_m) + u_i$$

which is nonlinear in both X and β, we can obtain the estimated parameters by nonlinear least squares (NLS) estimation.

The essential idea of NLS is the same as OLS, which is to minimize the sum of squared prediction mistakes. That is

$$\min_{b_1, \dots, b_m} S(b_1, \dots, b_m) = \sum_{i=1}^{n} [Y_i - f(X_1, \dots, X_k; b_1, \dots, b_m)]^2$$

The solution to this minimization problem is the nonlinear least squares estimators.