Chapter 5

Demand Estimation

Managerial Economics: Economic Tools for Today's Decision Makers, 4/e
By Paul Keat and Philip Young

Demand Estimation

- Regression Analysis
- The Coefficient of Determination
- Evaluating the Regression Coefficients
- Multiple Regression Analysis
- The Use of Regression Analysis to Forecast Demand
- Additional Topics
- Problems in the Use of Regression Analysis

Regression Analysis

Regression Analysis: A statistical technique for finding the best relationship between a dependent variable and selected independent variables.

- Simple regression – one independent variable
- Multiple regression – several independent variables

Regression Analysis

Dependent variable:

- depends on the value of other variables
- is of primary interest to researchers

Independent variables:

- used to explain the variation in the dependent variable

Regression Analysis

<u>Procedure</u>

1. Specify the regression model
2. Obtain data on the variables
3. Estimate the quantitative relationships
4. Test the statistical significance of the results
5. Use the results in decision making

Regression Analysis

Simple Regression

$$Y = a + bX + u$$

Y = dependent variable   X = independent variable

a = intercept   b = slope

u = random factor

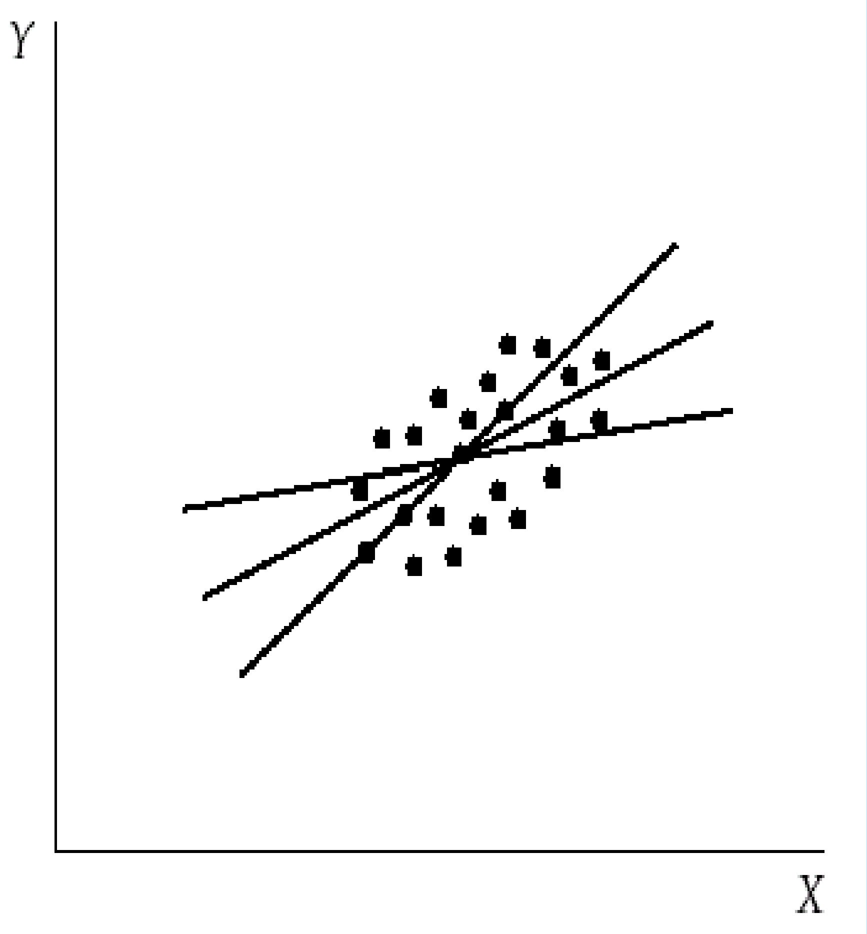
Regression Analysis

Data

- Cross-sectional data provide information on a group of entities at a given time.
- Time-series data provide information on one entity over time.

Regression Analysis

The estimation of the regression equation involves a search for the best linear relationship between the dependent and the independent variable.



Regression Analysis

<u>Method of ordinary least squares (OLS)</u>: A statistical method designed to fit a line through a scatter of points is such a way that the sum of the squared deviations of the points from the line is minimized.

Many software packages perform OLS estimation.

Regression Analysis

$$Y = a + bX$$

The intercept (a) and slope (b) of the regression line are referred to as the *parameters* or **coefficients of the regression equation**.

Coefficient of Determination

<u>Coefficient of determination (R^2)</u>: A measure indicating the percentage of the variation in the dependent variable accounted for by variations in the independent variables.

R^2 is a measure of the goodness of fit of the regression model.

Coefficient of Determination

Total sum of squares (TSS)

- Sum of the squared deviations of the sample values of Y from the mean of Y.
  - TSS = sum$(\bar{Y}_i - Y)^2$
    - Y_i = data (dependent variable)
    - $\bar{Y}$ = mean of the dependent variable
    - i = number of observations

Coefficient of Determination

Regression sum of squares (RSS)

- Sum of the squared deviations of the estimated values of Y from the mean of Y.
  - $RSS = sum(\hat{Y}_i - \bar{Y})^2$
    - $\hat{Y}_i$ = estimated value of Y
    - $\bar{Y}$ = mean of the dependent variable
    - i = number of observations

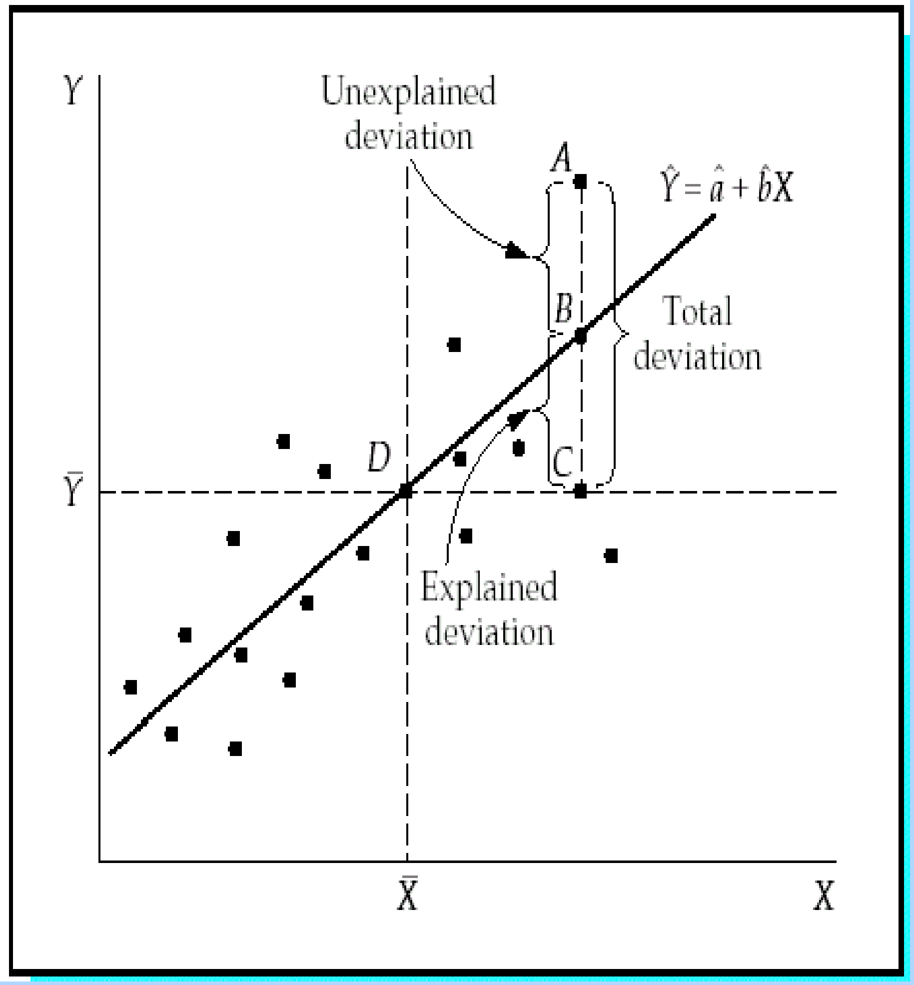
Coefficient of Determination

Error sum of squares (ESS)

- Sum of the squared deviations of the sample values of Y from the estimated values of Y.
  - $ESS = sum(Y_i - \hat{Y}_i)^2$
    - $\hat{Y}_i$ = estimated value of Y
    - Y_i = data (dependent variable)
    - i = number of observations

Coefficient of Determination

- TSS : see segment AC
- RSS: see segment BC
- ESS: see segment AB



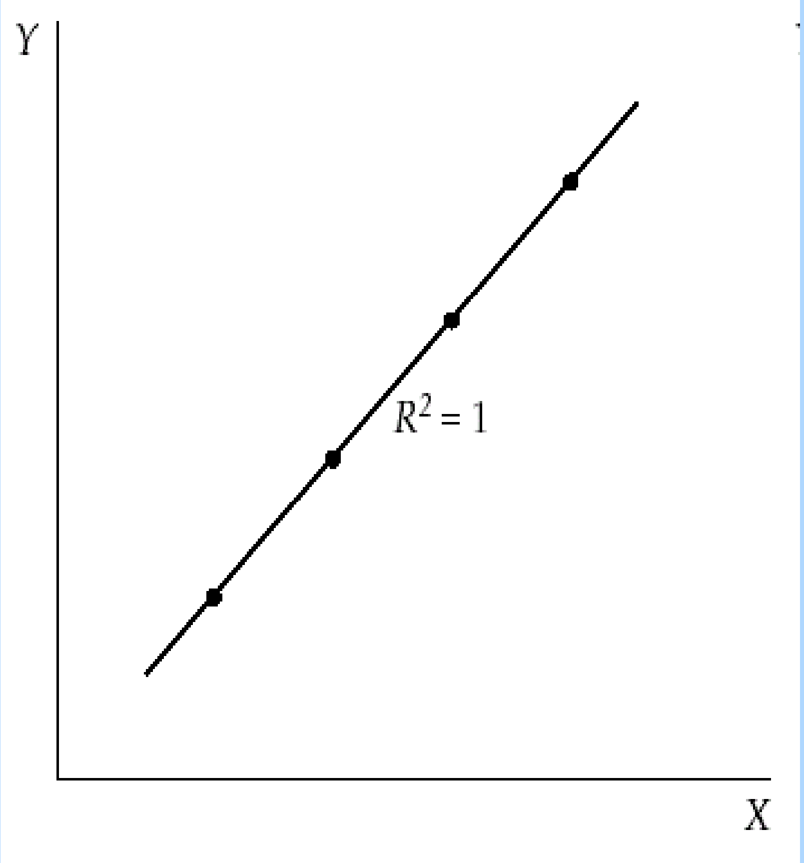
Coefficient of Determination

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

R^2 measures the proportion of the total deviation of Y from its mean which is explained by the regression model.

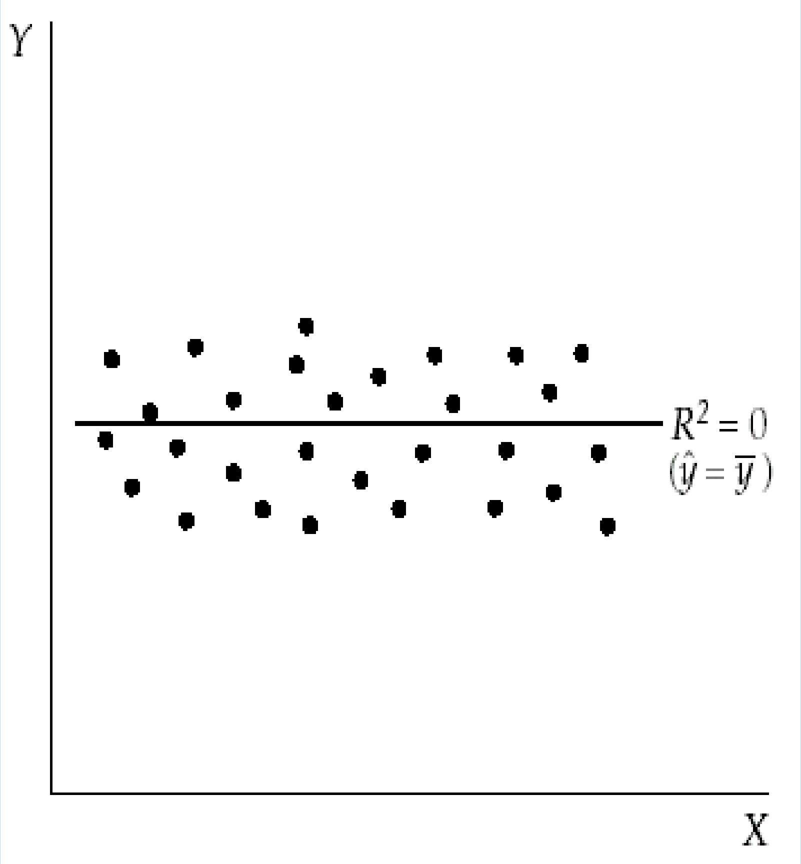
Coefficient of Determination

If $R^2 = 1$ the total deviation in Y from its mean is explained by the equation.



Coefficient of Determination

If $R^2 = 0$ the regression equation does not account for any of the variation of Y from its mean.



Coefficient of Determination

The closer R^2 is to unity, the greater the explanatory power of the regression equation.

An R^2 close to 0 indicates a regression equation will have very little explanatory power.

Coefficient of Determination

As additional independent variables are added, the regression equation will explain more of the variation in the dependent variable.

This leads to higher R^2 measures.

Coefficient of Determination

Adjusted coefficient of determination

$$\overline{R}^2 = R^2 - \frac{k}{n - k - 1}(1 - R^2)$$

k = number of independent variables

n = sample size

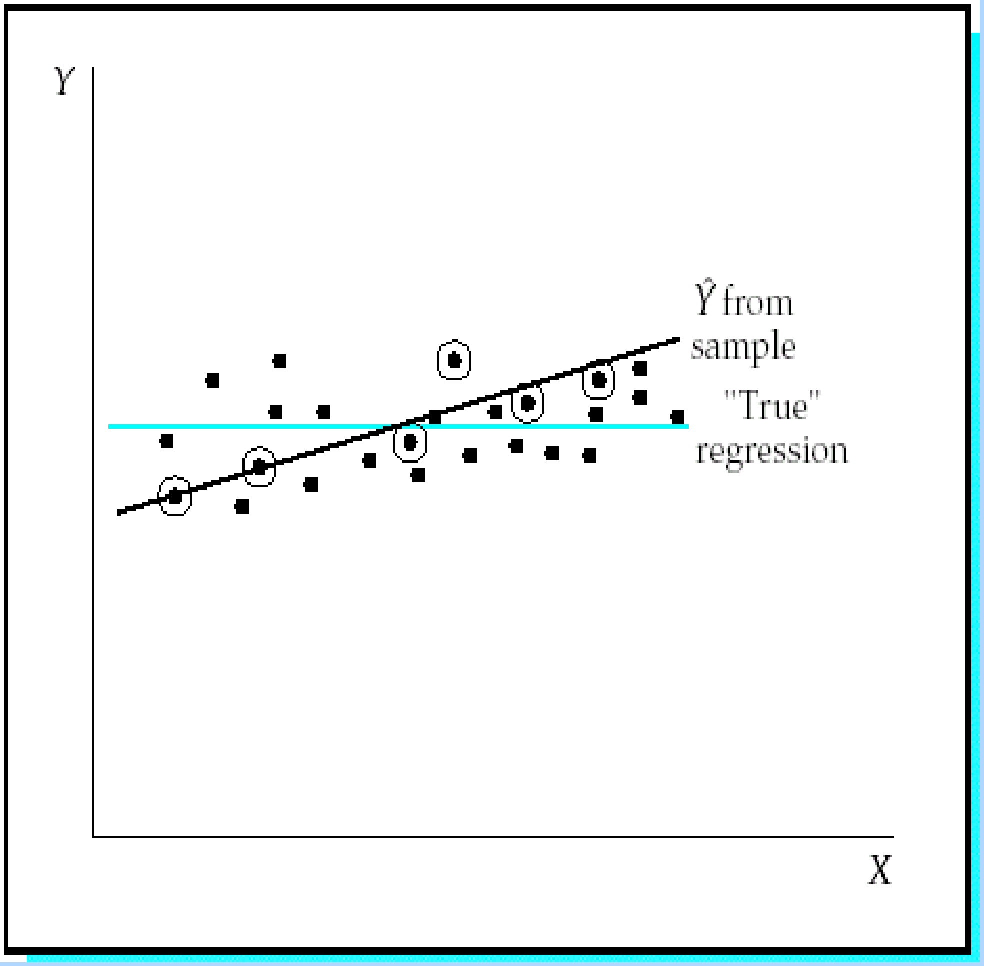
Evaluating the Regression Coefficients

In most cases, a sample from the population is used rather than the entire population.

It becomes necessary to make inferences about the population based on the sample and to make a judgment about how good these inferences are.

Evaluating the Regression Coefficients

An OLS regression line fitted through the sample points may differ from the true (but unknown) regression line.



Evaluating the Regression Coefficients

How confident can a researcher be about the extent to which the regression equation for the sample truly represents the unknown regression equation for the population?

Evaluating the Regression Coefficients

Each random sample from the population generates its own intercept and slope coefficients.

To determine whether b (or a) is statistically different from 0 we conduct a t-test.

Evaluating the Regression Coefficients

- Two-tail test

Null Hypothesis

$$H_0 : b = 0$$

Alternative Hypothesis

$$H_a : b \neq 0$$

- One-tail test

Null Hypothesis

$$H_0 : b > 0 \text{ (or } b < 0)$$

Alternative Hypothesis

$$H_a : b < 0 \text{ (or } b > 0)$$

Evaluating the Regression Coefficients

Test statistic

$$t = \frac{b - E(b)}{SE_b}$$

b = estimated coefficient

$E(b) = b = 0$ (Null hypothesis)

SE_b = standard error of the coefficient

Evaluating the Regression Coefficients

Critical t-value depends on:

- Degrees of freedom (d.f. = n – k – 1)
- One or two-tailed test
- Level of significance

Use a t-table to determine the critical t-value.

Evaluating the Regression Coefficients

Compare the t-value with the critical value.

Reject the null hypothesis if the absolute value of the test statistic is greater than or equal to the critical t-value.

Fail to reject the null hypothesis if the absolute value of the test statistic is less than the critical t-value.

Multiple Regression Analysis

In **multiple regression analysis** the coefficients indicate the change in the dependent variable assuming the values of the other variables are unchanged.

Multiple Regression Analysis

An additional test of statistical significance is called the F-test.

The F-test measures the statistical significance of the entire regression equation rather than each individual coefficient.

Multiple Regression Analysis

Null Hypothesis

$H_0: b_1 = b_2 = b_3 = \ldots = b_k = 0$

No relationship exists between the dependent variable and the k independent variables for the population.

Multiple Regression Analysis

- F-test statistic

$$F = \frac{\left(R^2 / k \right)}{\left(1 - R^2\right) / \left(n - k - 1\right)}$$

Multiple Regression Analysis

Critical F-value (F*) depends on:

- Numerator degrees of freedom
  - (n.d.f. = k)
- Denominator degrees of freedom
  - (d.d.f = n-k-1)
- Level of significance

Use a F-table to determine the critical F-value.

Multiple Regression Analysis

Compare the F-value with the critical value.

- If $F > F^*$
- Reject Null Hypothesis
- The entire regression model accounts for a statistically significant portion of the variation in the dependent variable.

Multiple Regression Analysis

Compare the F-value with the critical value.

- If $F < F^*$
- Fail to reject Null Hypothesis
- There is no statistically significant relationship between the dependent variable and all of the independent variables.

The Use of Regression Analysis to Forecast Demand

Forecast of dependent variable

$$\bar{Y} \pm t_{n-k-1} SEE$$

SEE = Standard error of the estimate

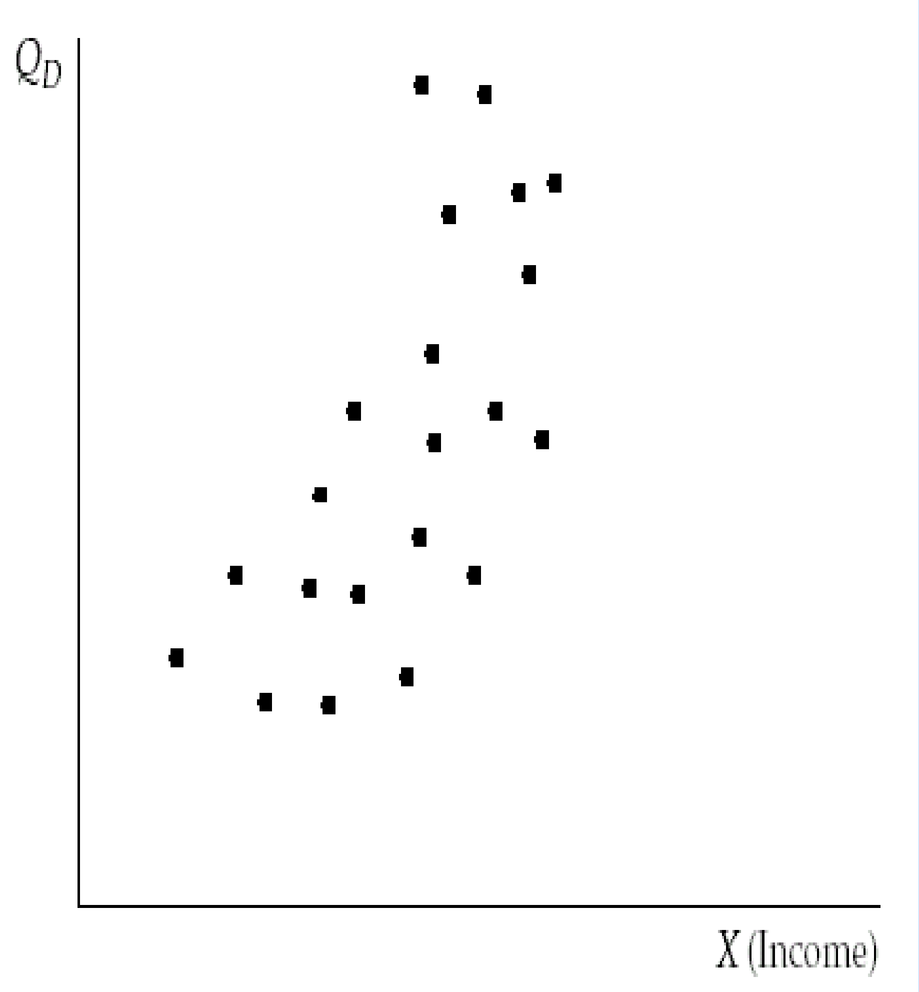
Additional Topics

Proxy variable: an alternative variable used in a regression when direct information in not available

Dummy variable: a binary variable created to represent a non-quantitative factor.

Additional Topics

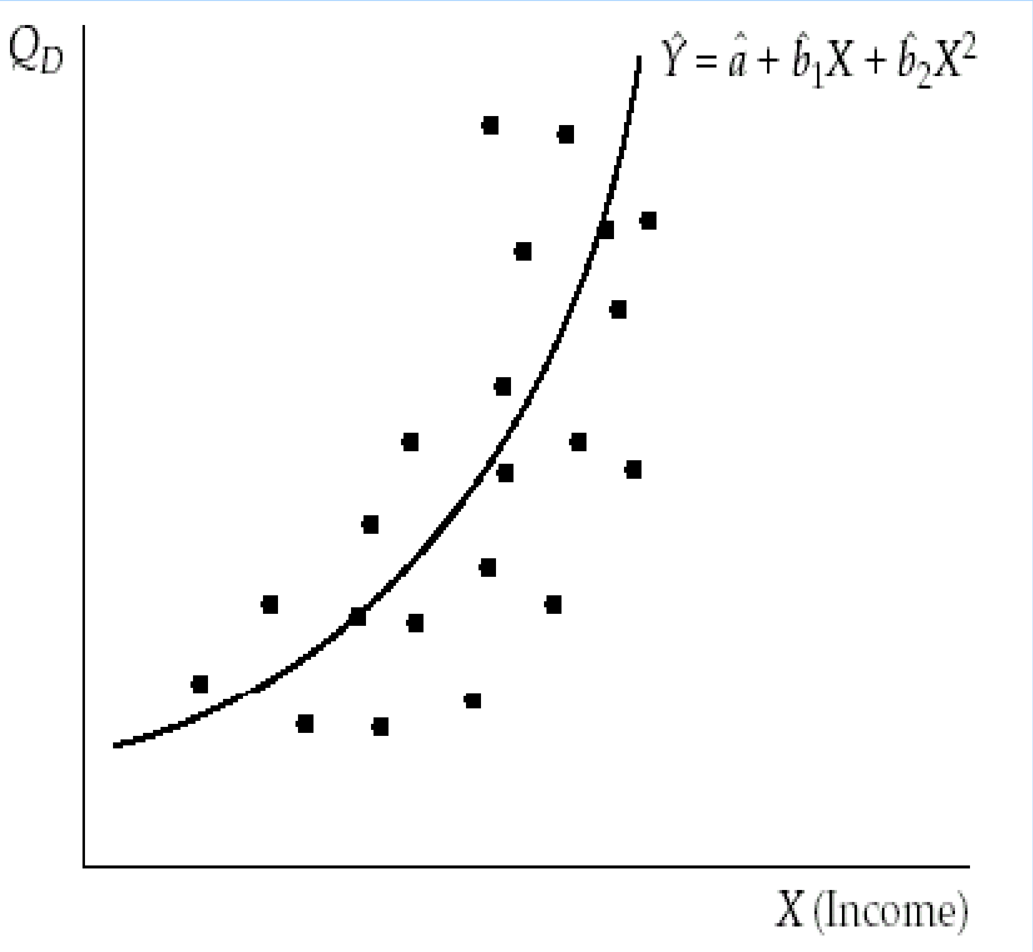
The relationship between the dependent and independent variables may be nonlinear.



Additional Topics

We could specify the regression model as quadratic regression model.

$$Y = a + b_1 x + b_2 x^2$$



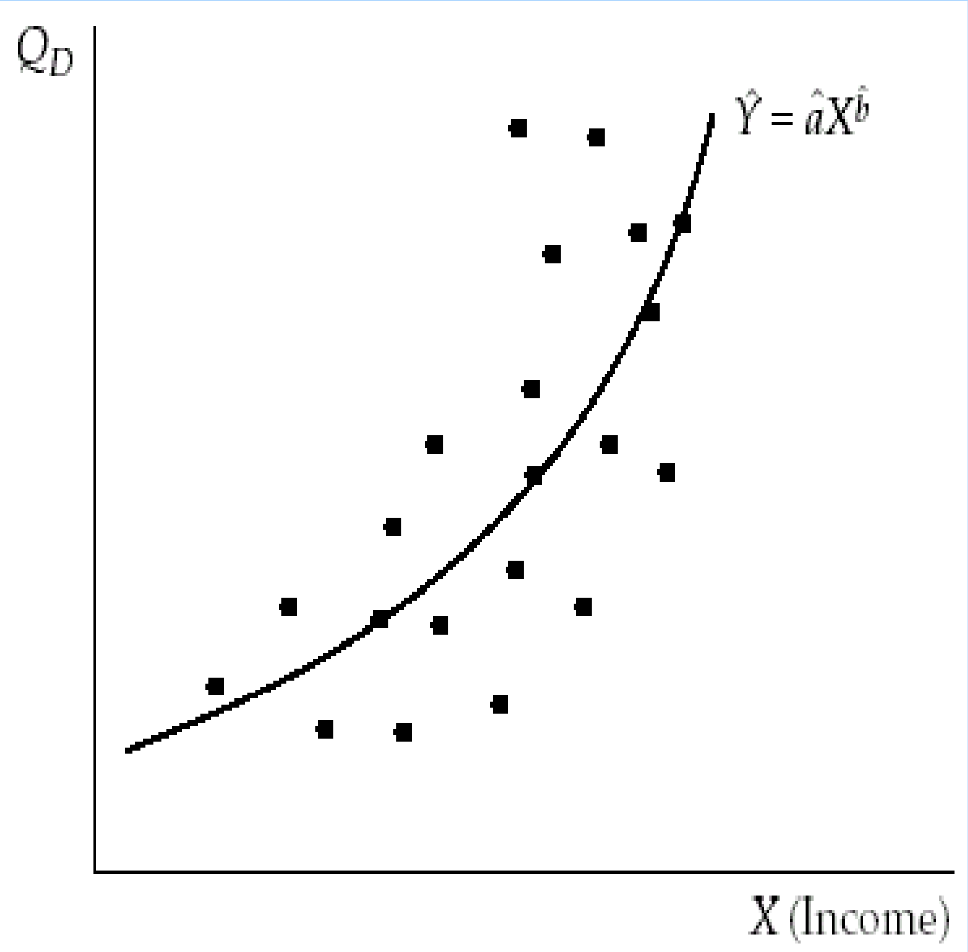
Additional Topics

We could also specify the regression model as power function.

$$Y = ax^b$$

or

$$\log Q_d = \log a + b(\log X)$$



Problems

The estimation of demand may produce biased results due to simultaneous shifting of supply and demand curves.

This is referred to as the **identification problem**.

Advanced estimation techniques, such as two-stage least squares, are used to correct this problem.

Problems

If two independent variables are closely associated, it becomes difficult to separate the effects of each on the dependent variable.

If the regression passes the F-test but fails the t-test for each coefficient, **multicollinearity** exists.

A standard remedy is to drop one of the closely related independent variables from the regression.

Problems

Autocorrelation occurs when the dependent variable deviates from the regression line in a systematic way.

The *Durbin-Watson* statistic is used to identify the presence of autocorrelation.

To correct autocorrelation consider:

- Transforming the data into a different order of magnitude.
- Introducing leading or lagging data