# Chapter 5
# The Multiple Regression Model

# Chapter Contents

- 5.1 Introduction
- 5.2 Estimating the Parameters of the Multiple Regression Model
- 5.3 Finite Sample Properties of the Least Squares Estimator
- 5.4 Interval Estimation
- 5.5 Hypothesis Testing
- 5.6 Nonlinear Relationships
- 5.7 Large Sample Properties of the Least Squares Estimator

# 5.1 Introduction

- When we turn an economic model with more than one explanatory variable into its corresponding econometric model, we refer to it as a multiple regression model
- Most of the results we developed for the simple regression model in Chapters 2–4 can be extended naturally to this general case
- There are slight changes in the interpretation of the $\beta$ parameters, the degrees of freedom for the t-distribution will change
- We will need to modify the assumption concerning the characteristics of the explanatory (x) variables

# 5.1.1 The Economic Model 1 of 2

- Let's set up an economic model in which sales revenue depends on one or more explanatory variables
  - We initially hypothesize that sales revenue is linearly related to price and advertising expenditure
  - The economic model is:
  - (5.1) $$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT$$

# 5.1.1 The Economic Model 2 of 2

- In most economic models there are two or more explanatory variables
  - When we turn an economic model with more than one explanatory variable into its corresponding econometric model, we refer to it as a **multiple regression model**
  - Most of the results we developed for the simple regression model can be extended naturally to this general case

# 5.1.2 The Econometric Model 1 of 2

- To allow for a difference between observable sales revenue and the expected value of sales revenue, we add a random error term, e = SALES - E(SALES)
  - (5.2) $SALES = E\left(SALES\right) + e = \beta_1 + \beta_2 PRICE + \beta_3 \, ADVERT + e$
- Equation (5.5) is the conditional mean or conditional expectation of SALES given PRICE and ADVERT
- It is also known as the multiple regression function or simply the regression function
  - (5.5) $E(SALES|PRICE, ADVERT) = \beta 1 + \beta 2 PRICE + \beta 3 ADVERT$

# 5.1.2 The Econometric Model 2 of 2

- $\beta_2$ is the change in monthly sales *SALES* ($1000) when the price index *PRICE* is increased by one unit ($1), and advertising expenditure *ADVERT* is held constant
- Similarly, $\beta_3$ is the change in monthly sales *SALES* ($1000) when the advertising expenditure is increased by one unit ($1000), and the price index *PRICE* is held constant

$$\beta_2 = \frac{\Delta SALES}{\Delta PRICE \left( ADVERT \text{ held constant} \right)} = \frac{\partial SALES}{\partial PRICE}$$

$$\beta_3 = \frac{\Delta SALES}{\Delta ADVERT \left( PRICE \text{ held constant} \right)} = \frac{\partial SALES}{\partial ADVERT}$$

# Figure 5.1 The multiple regression plane

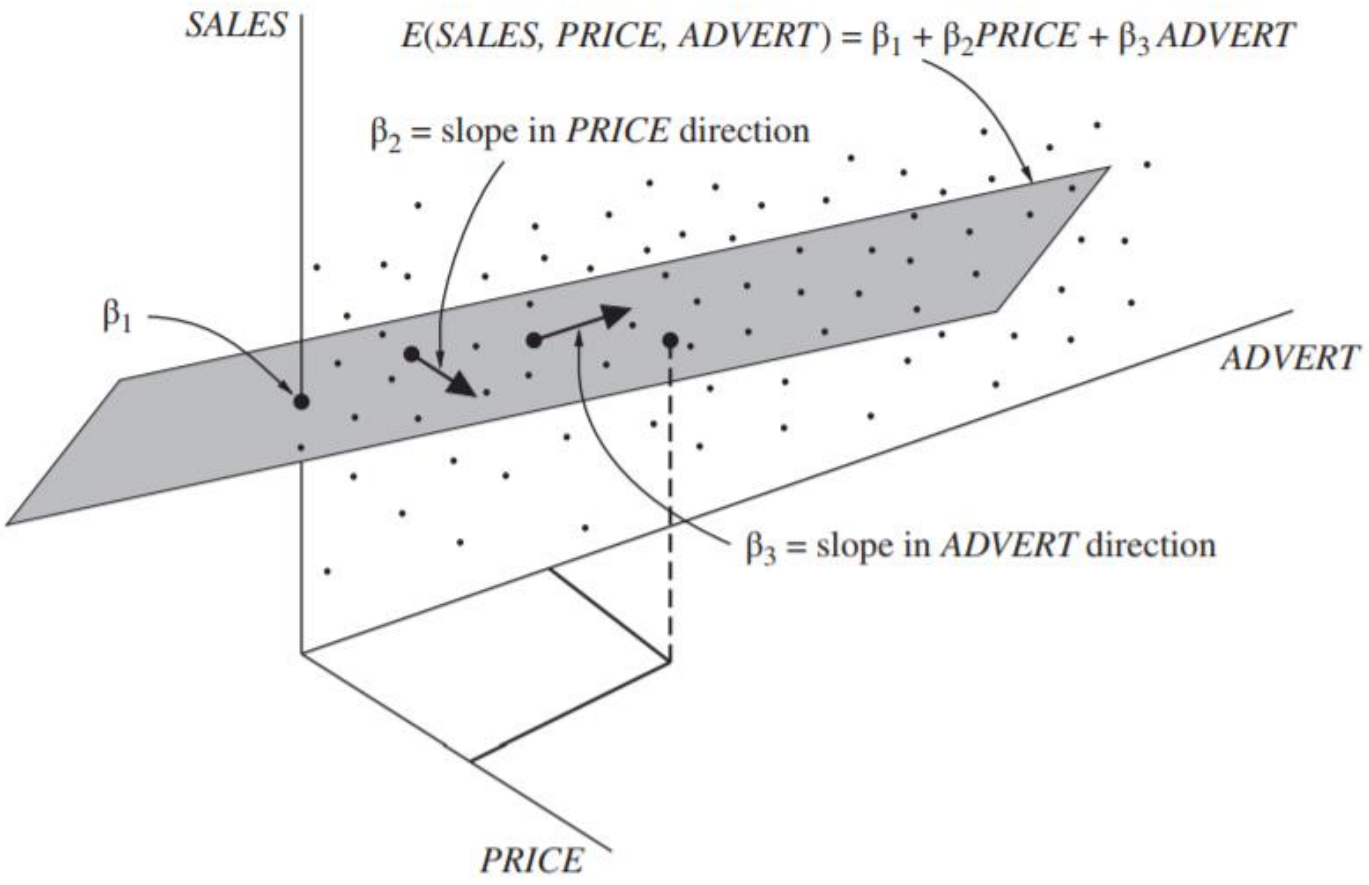


**FIGURE 5.1** **The multiple regression plane.**

# 5.1.3 The General Model

- Let: $y_i = SALES_i,\ x_{i2} = PRICE_i,\ \ x_{i3} = ADVERT_i$
- Then (5.6) $y_i = \beta_1 + \beta_2 x_{i2_i} + \beta_3 x_{i3} + e_i$
- We can think of the first term on the right-hand side of the equation as $\beta_1\ x_{i1}$ where $x_{i1} = 1$, that is, $x_{i1}$ is equal to 1 for all observations; it is called the constant term
- There are many multiple regression models where we have more than two explanatory variables

# 5.1.4 Assumptions of the Multiple Regression Model 1 of 2

- MR1: Econometric Model Observations on $(y_i, x_i) = (y_i, x_{i2}, x_{i3}, \dots x_{iK})$ satisfy the population relationship $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{iK} + e_i$
- MR2: Strict Exogeneity The conditional expectation of the random error $e_i$ , given all explanatory variable observations $\mathbf{X} = \{x_i, i\ 1, 2, \dots, N\}, is\ 0. E(e_i|X) = 0$
- MR3: Conditional Homoskedasticity The variance of the error term, conditional on X, is a constant var$(e_i|X) = \sigma^2$

# 5.1.4 Assumptions of the Multiple Regression Model 2 of 2

- MR4: Conditionally Uncorrelated Errors The covariance between different error terms $e_i$ and $e_j$ , conditional on X, is zero $\text{cov}(e_i , e_j |X ) = 0$ for $i \neq j$
- MR5: No Exact Linear Relationship Exists Between the Explanatory Variables It is not possible to express one of the explanatory variables as an exact linear function of the others
- MR6: Error Normality (optional) Conditional on X, the errors are normally distributed

# 5.2 Estimating the Parameters of the Multiple Regression Model

- We will discuss estimation in the context of the model in (5.6), which we repeat here for convenience, with $i$ denoting the $i$th observation

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$$

- This model is simpler than the full model, yet all the results we present carry over to the general case with only minor modifications

# 5.2.1 Least Squares Estimation Procedure 1 of 2

- Mathematically we minimize the sum of squares function $S(\beta_1, \beta_2, \beta_3)$, which is a function of the unknown parameters, given the data:

- (5.9)

$$S(\beta_1, \beta_2, \beta_3) = \sum_{i=1}^{N}\left(y_i - E(y_i)\right)^2$$

$$= \sum_{i=1}^{N}\left(y_i - \beta_1 - \beta_2 x_{i2} - \beta_3 x_{i3}\right)^2$$

# 5.2.1 Least Squares Estimation Procedure 2 of 2

- Formulas for $b_1$, $b_2$, and $b_3$, obtained by minimizing (5.9), are estimation procedures, which are called the **least squares estimators** of the unknown parameters
  - In general, since their values are not known until the data are observed and the estimates calculated, the least squares estimators are random variables
  - These least squares estimators and estimates are also referred to as ordinary least squares estimators and estimates, abbreviated OLS, to distinguish them from other estimators

# 5.2.2 Estimating the Error Variance $\sigma^2$ 1 of 2

- We need to estimate the error variance, $\sigma^2$
- Recall that: $\sigma^2 = \text{var}(e_i) = E\left(e_i^2\right)$
- But, the squared errors are unobservable, so we develop an estimator for $\sigma^2$ based on the squares of the least squares residuals:

$$\hat{e}_i = y_i - y_i = y_i - \left(b_1 + b_2 x_{i2} + b_3 x_{i3}\right)$$

# 5.2.2 Estimating the Error Variance $\sigma^2$ 2 of 2

- An estimator for $\sigma^2$ that uses the information from $\widehat{e_i^2}$ and has good statistical properties is
  - (5.11) $$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{e}_i^2}{N-K}$$
- Where $K$ is the number of β parameters being estimated in the multiple regression model.

# 5.2.3 Measuring Goodness-of-Fit 1 of 2

- For the simple regression model studied in Chapter 4, we introduced $R^2$ as a measure of the proportion of variation in the dependent variable that is explained by variation in the explanatory variable
- We talk of the proportion of variation in the dependent variable explained by all the explanatory variables included in the model
- The coefficient of determination is
- (5.12) $R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{N}(\hat{y}i - \bar{y})^2}{\sum_{i=1}^{N}(yi - \bar{y})^2} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{N}(\widehat{e_i^2})}{\sum_{i=1}^{N}(yi - \bar{y})^2}$

# 5.2.3 Measuring Goodness-of-Fit 2 of 2

- Where SSR is the variation in y "explained" by the model (sum of squares due to regression regression)
- SST is the total variation in y about its mean (sum of squares, total)
- SSE is the sum of squared least squares residuals (errors) and is that part of the variation in y that is not explained by the model
- The coefficient of determination is also viewed as a measure of the predictive ability of the model over the sample period, or as a measure of how well the estimated regression fits the data

# 5.2.4 Frisch–Waugh–Lovell (FWL) Theorem 1 of 2

- It helps understand in a multiple regression the interpretation of a coefficient estimate, all other variables held constant
- To illustrate this result, we use the sales equation $SALES_i = \beta_1 + \beta_2 PRICE_i + \beta_3 ADVERT_i + e_i$ and carry out the following steps:

1. Estimate the simple regression $SALES_i = a_1 + a_2\ PRICE_i$ + error using the least squares estimator and save the least squares residuals
2. Estimate the simple regression $ADVERT_i = c_1 + c_2\ PRICE_i$ + error using the least squares estimator and save the least squares residual

# 5.2.4 Frisch–Waugh–Lovell (FWL) Theorem 2 of 2

3. Estimate the simple regression $\widetilde{SALES}_i = \beta_3 \, \widetilde{ADVERT}_i + \tilde{e}_i$ with no constant term.

   The estimate of $\beta_3$ is b3 = 1.8626. This estimate is the same as that reported from

   the full regression in Table 5.2

4. Compute the least squares residuals from step 3, $\tilde{e}_i = \widetilde{SALES}_i - \text{b3} \, \widetilde{ADVERT}_i$ .

   Compare these residuals to those from the complete model

# 5.3 Finite Sample Properties of the Least Squares Estimator

- For the multiple regression model, if assumptions MR1–MR5 hold, then the least squares estimators are the Best Linear Unbiased Estimators (*BLUE*) of the parameters in the multiple regression model
- If the errors are not normally distributed, then the least squares estimators are approximately normally distributed in large samples
- These various properties—BLUE and the use of the t-distribution for interval estimation and hypothesis testing—are finite sample properties. As long as $N > K$, they hold irrespective of the sample size N

# 5.3.1 The Variances and Covariances of the Least Squares Estimators 1 of 3

- We can show that:
  - (5.13) $var(b_2|x) = \frac{\sigma^2}{(1-r_{23}^2)\sum_{i=1}^{N}(x_{i2}-\overline{\overline{x_2}})^2}$
- Where
  - $r_{23}^2 = \frac{\sum(x_{i2}-\overline{\overline{x_2}})(x_{i3}-\overline{\overline{x_3}})}{\sqrt{\sum(x_{i2}-\overline{\overline{x_2}})^2\sum(x_{i3}-\overline{\overline{x_3}})^2}}$

# 5.3.1 The Variances and Covariances of the Least Squares Estimators 2 of 3

- We can see that:

1. Larger error variances $\sigma^2$ lead to larger variances of the least squares estimators
2. Larger sample sizes $N$ imply smaller variances of the least squares estimators
3. More variation in an explanatory variable around its mean, leads to a smaller variance of the least squares estimator
4. A larger correlation between $x_2$ and $x_3$ leads to a larger variance of $b_2$

# 5.3.1 The Variances and Covariances of the Least Squares Estimators 3 of 3

- It is customary to arrange the estimated variances and covariances of the least squares estimators in a square array, which is called a matrix
- This matrix has variances on its diagonal and covariances in the off-diagonal positions. It is called a variance–covariance matrix

$$\operatorname{cov}(b_1, b_2, b_3) = \begin{bmatrix} \operatorname{var}(b_1) & \operatorname{cov}(b_1, b_2) & \operatorname{cov}(b_1, b_3) \\ \operatorname{cov}(b_1, b_2) & \operatorname{var}(b_2) & \operatorname{cov}(b_2, b_3) \\ \operatorname{cov}(b_1, b_3) & \operatorname{cov}(b_2, b_3) & \operatorname{var}(b_3) \end{bmatrix}$$

# 5.3.2 The Distribution of the Least Squares Estimators 1 of 4

- Consider the general for of a multiple regression model:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_K x_{iK} + e_i$$

  - If we add assumption MR6, that the random errors $e_i$ have normal probability distributions, then the dependent variable $y_i$ is normally distributed:
  - $(e_i|X) \sim N(0, \sigma^2)$

# 5.3.2 The Distribution of the Least Squares Estimators 2 of 4

- Since the least squares estimators are linear functions of dependent variables, it follows that the least squares estimators are also normally distributed:
- $(b_k|X) \sim N(\beta_k, var(b_k|X))$
- We can now form the standard normal variable $Z$:
  - (5.14) $z = \frac{b_k - \beta_k}{\sqrt{\text{var}(b_k)}} \sim N(0,1), \text{ for } k = 1, 2, \ldots, K$

# 5.3.2 The Distribution of the Least Squares Estimators 3 of 4

- We can form a linear combination of the coefficients as:

$$\lambda = c_1\beta_1 + c_2\beta_2 + \ldots + c_K\beta_K = \sum_{k=1}^{K} c_k\beta_k$$

- And then we have
  - (5.16) $$t = \frac{\hat{\lambda} - \lambda}{se\left(\hat{\lambda}\right)} = \frac{\sum c_k b_k - \sum c_k \beta_k}{\mathrm{se}\left(\sum c_k b_k\right)} \sim t_{(N-K)}$$

# 5.3.2 The Distribution of the Least Squares Estimators 4 of 4

- What happens if the errors are not normally distributed?
  - Then the least squares estimator will not be normally distributed and (5.11), (5.12), and (5.13) will not hold exactly
  - They will, however, be approximately true in large samples
  - Thus, having errors that are not normally distributed does not stop us from using (5.12) and (5.13), but it does mean we have to be cautious if the sample size is not large
  - A test for normally distributed errors was given in Chapter 4.3.5

# 5.4.1 Interval Estimation for a Single Coefficient 1 of 3

- Suppose we are interested in finding a 95% interval estimate for $\beta_2$, the response of average sales revenue to a change in price at Big Andy's Burger Barn
- The first step is to find a value from the t(72)-distribution, call it $t_c$, such that
  - (5.17) $P(-t_c < t_{(72)} < t_c) = .95$
- Using $t_c$ = 1.993, we can rewrite (5.17) as:

$$P\left(-1.993 \le \frac{b_2 - \beta_2}{\text{se}(b_2)} \le 1.993\right) = .95$$

# 5.4.1 Interval Estimation for a Single Coefficient 2 of 3

- The interval endpoints
  - (5.18) $\left[b_2 - 1.993 \times \text{se}(b_2),\ b_2 + 1.993 \times \text{se}(b_2)\right]$
- If this interval estimator is used in many samples from the population, then 95% of them will contain the true parameter $\beta_2$
- Before the data are collected, we have confidence in the interval estimation procedure (estimator) because of its performance over all possible samples

# 5.4.1 Interval Estimation for a Single Coefficient 3 of 3

- In general, if an interval estimate is uninformative because it is too wide, there is nothing immediate that can be done
- A narrower interval can only be obtained by reducing the variance of the estimator
- We cannot say, in general, what constitutes an interval that is too wide, or too uninformative. It depends on the context of the problem being investigated
- To give a general expression for an interval estimate, we need to recognize that the critical value $t_c$ will depend on the degree of confidence specified for the interval estimate and the number of degrees of freedom

# 5.4.2 Interval Estimation for a Linear Combination of Coefficients 1 of 3

- **Example 5.7 Interval Estimate for a Change in Sales**
- Suppose Big Andy wants to increase advertising expenditure by \$800 and drop the price by 40 cents.
- Then the change in expected sales is:

$$
\begin{aligned}
\lambda &= E\left(SALES_1\right) - E\left(SALES_0\right) \\
&= \left[\beta_1 + \beta_2\left(PRICE_0 - 0.4\right) + \beta_3\left(ADVERT_0 + 0.8\right)\right] \\
&\quad -\left[\beta_1 + \beta_2 PRICE_0 + \beta_3 ADVERT_0\right] \\
&= -0.4\beta_2 + 0.8\beta_3
\end{aligned}
$$

# 5.4.2 Interval Estimation for a Linear Combination of Coefficients 2 of 3

- A point estimate would be:

$$\hat{\lambda} = -0.4\text{b}_2 + 0.8\text{b}_3 = -0.4 \times (-7.9079) + 0.8 \times 1.8626$$
$$= 4.6532$$

- A 90% interval would be:

$$\left(\hat{\lambda} - t_c \times se(\lambda), \hat{\lambda} - t_c \times se(\lambda)\right)$$
$$= \Big(\left(-0.4b_2 + 0.8b_3\right) - t_c \times se\left(-0.4b_2 + 0.8b_3\right),$$
$$\left(-0.4b_2 + 0.8b_3\right) + t_c \times se\left(-0.4b_2 + 0.8b_3\right)\Big)$$

# 5.4.2 Interval Estimation for a Linear Combination of Coefficients 3 of 3

- Thus, a 90% interval is:
  - $(4.6532 - 1.666 \times 0.7096, 4.6532 + 1.666 \times 0.7096) = (3.471, 5.835)$
- We estimate, with 90% confidence, that the expected increase in sales from Big Andy’s strategy will lie between \$3471 and \$5835.

# 5.5 Hypothesis Testing

- Step-By-Step Procedure for Testing Hypotheses

1. Determine the null and alternative hypotheses

2. Specify the test statistic and its distribution if the null hypothesis is true

3. Select $\alpha$ and determine the rejection region

4. Calculate the sample value of the test statistic and, if desired, the p-value

5. State your conclusion

# 5.5.1 Testing the Significance of a Single Coefficient 1 of 2

- When we set up a multiple regression model, we do so because we believe the explanatory variables influence the dependent variable y
  - If a given explanatory variable, say $x_k$, has no bearing on $y$, then $\beta_k = 0$
  - Testing this null hypothesis is sometimes called a **test of significance** for the explanatory variable $x_k$

# 5.5.1 Testing the Significance of a Single Coefficient 2 of 2

- Null hypothesis: $H_0\colon\ \beta_k = 0$

- Alternative hypothesis: $H_1\colon\ \beta_k \neq 0$

- Test statistic:

$$t = \frac{b_k}{\operatorname{se}(b_k)} \sim t_{(N-K)}$$

- $t$ values for a test with level of significance α: $t_c = t_{(1-\alpha/2,\, N-K)}$ and $-t_c = t_{(\alpha/2,\, N-K)}$

# 5.5.2 One-Tail Hypothesis Testing for a Single Coefficient 1 of 3

- **Example 5.10 Testing for Elastic Demand**
- $\beta_2 \geq 0$: a decrease in price leads to a change in sales revenue that is zero or negative (demand is price-inelastic or has an elasticity of unity)
- $\beta_2 < 0$: a decrease in price leads to an increase in sales revenue (demand is price-elastic)

# 5.5.2 One-Tail Hypothesis Testing for a Single Coefficient 2 of 3

- The null and alternative hypotheses are:

$$H_0 : \beta_2 \geq 0 \quad \text{(demand is unit-elastic or inelastic)}$$

$$H_1 : \beta_2 < 0 \quad \text{(demand is elastic)}$$

- The test statistic, if the null hypothesis is true, is: $t = b_2 / \text{se}(b_2) \sim t_{(N-K)}$

- At a 5% significance level, we reject $H_0$ if $t \leq -1.666$ or if the $p$-value $\leq 0.05$

# 5.5.2 One-Tail Hypothesis Testing for a Single Coefficient 3 of 3

- The test statistic is:

$$t = \frac{b_2}{se(b_2)} = \frac{-7.908}{1.096} = -7.215$$

- and the $p$-value is:

$$P\left(t_{(72)} < -7.215\right) = 0.000$$

- Since -7.215 < 1.666, we reject $H_0$: $\beta_2 \geq 0$ and conclude that $H_0$: $\beta_2 < 0$ (demand is elastic)

# 5.5.3 Hypothesis Testing for a Linear Combination of Coefficients 1 of 2

- **Example 5.12 Testing the Effect of Changes in Price and Advertising**
- The null and alternative hypotheses are:

$$H_0: -0.2\beta_2 - 0.5\beta_3 \leq 0 \qquad \text{(marketer's claim is not correct)}$$

$$H_1: -0.2\beta_2 - 0.5\beta_3 > 0 \qquad \text{(marketer's claim is correct)}$$

- The test statistic, if the null hypothesis is true, is:

$$t = \frac{-0.2b_2 - 0.5b_3}{\text{se}(-0.2b_2 - 0.5b_3)} \sim t_{(72)}$$

# 5.5.3 Hypothesis Testing for a Linear Combination of Coefficients 2 of 2

- At a 5% significance level, we reject $H_0$ if $t \geq 1.666$ or if the $p$-value $\leq 0.05$
- The test statistic is:

$$t = \frac{-0.2b_2 - 0.5b_3}{\text{se}\left(-0.2b_2 - 0.5b_3\right)} = \frac{1.58158 - 0.9319}{0.4010} = 1.622$$

- The $p$-value is: $P\left(t_{(72)} > 1.622\right) = 0.055$
- Since 1.622 < 1.666, we do not reject $H_0$

# 5.6 Nonlinear Relationships

- We have studied the multiple regression model: $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_K x_K + e$
- Sometimes we are interested in **polynomial** equations such as the quadratic
  - $y = \beta_1 + \beta_2 x + \beta_3 x^2 + e$
  - Or the cubic
  - $y = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3 + e$

# Example 5.13 Cost and Product Curves 1 of 3

- Consider the average cost equation:

$$AC = \beta_1 + \beta_2 Q + \beta_3 Q^2 + e$$

- And the total cost function:

$$TC = \alpha_1 + \alpha_2 Q + \alpha_3 Q^2 + \alpha_4 Q^3 + e$$

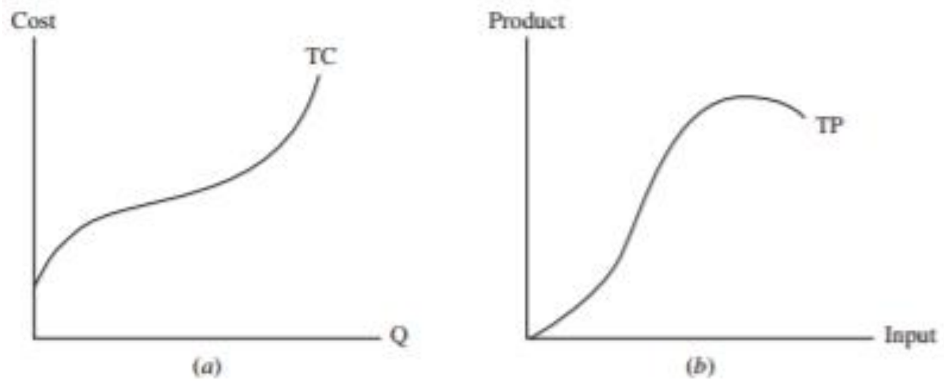


FIGURE 5.2 (a) Total cost curve and (b) total product curve.

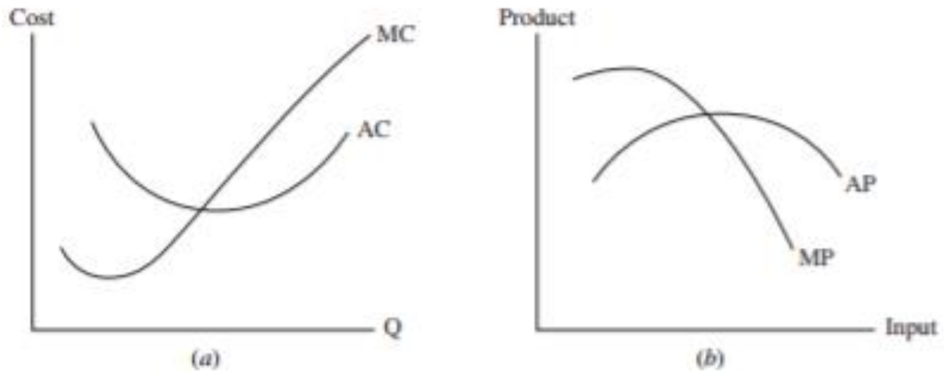


FIGURE 5.3 Average and marginal (a) cost curves and (b) product curves.

# Example 5.13 Cost and Product Curves 2 of 3

- For the general polynomial function:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_p x^p$$

- The slope is

$$\frac{dy}{dx} = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + pa_p x^{p-1}$$

- Evaluated at a particular value, $x = x_0$ , the slope is:

$$\left.\frac{dy}{dx}\right|_{x=x_0} = a_1 + 2a_2 x_0 + 3a_3 x_0^2 + \cdots + pa_p x_0^{p-1}$$

# Example 5.13 Cost and Product Curves 3 of 3

- The slope of the average cost curve is

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

- For this *U*-shaped curve, we expect $\beta_2 < 0$ and $\beta_3 > 0$

- The slope of the total cost curve (5.21), which is the marginal cost, is

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

- For a *U*-shaped marginal cost curve, we expect the parameter signs to be $\alpha_2 > 0$, $\alpha_3 < 0$, and $\alpha_4 > 0$

# 5.7 Large Sample Properties of the Least Squares Estimator

- Large sample approximate properties are known as asymptotic properties
- A question students always ask and instructors always evade is “how large does the sample have to be
- The answer depends on the model, the estimator, and the function of parameters that is of interest. Sometimes $N = 30$ is adequate; sometimes $N = 1000$ or larger could be necessary

# 5.7.1 Consistency 1 of 2

- When choosing econometric estimators, we do so with the objective in mind of obtaining an estimate that is close to the true but unknown parameter with high probability

- An estimator is said to be consistent if this probability converges to 1 as the sample size $N \rightarrow \infty$. Or, using the concept of a limit, the estimator b2 is consistent if

- $\lim_{n \to \infty} P\,(\beta_2 - \varepsilon \leq b_1 \leq \beta_2 + \varepsilon) = 1$

# 5.7.1 Consistency 2 of 2

- As the sample size increases the probability density function (*pdf*) becomes narrower
- As the sample size changes, the center of the pdfs remains at $\beta_2$. However, as the sample size N gets larger, the variance of the estimator $b_2$ becomes smaller
- The probability that $b_2$ falls in the interval $\beta_2 - \varepsilon \leq b_2 \leq \beta_2 + \varepsilon$ is the area under the pdf between these limits
- As the sample size increases, the probability of $b_2$ falling within the limits increases toward 1

# Figure 5.5 An illustration of consistency.

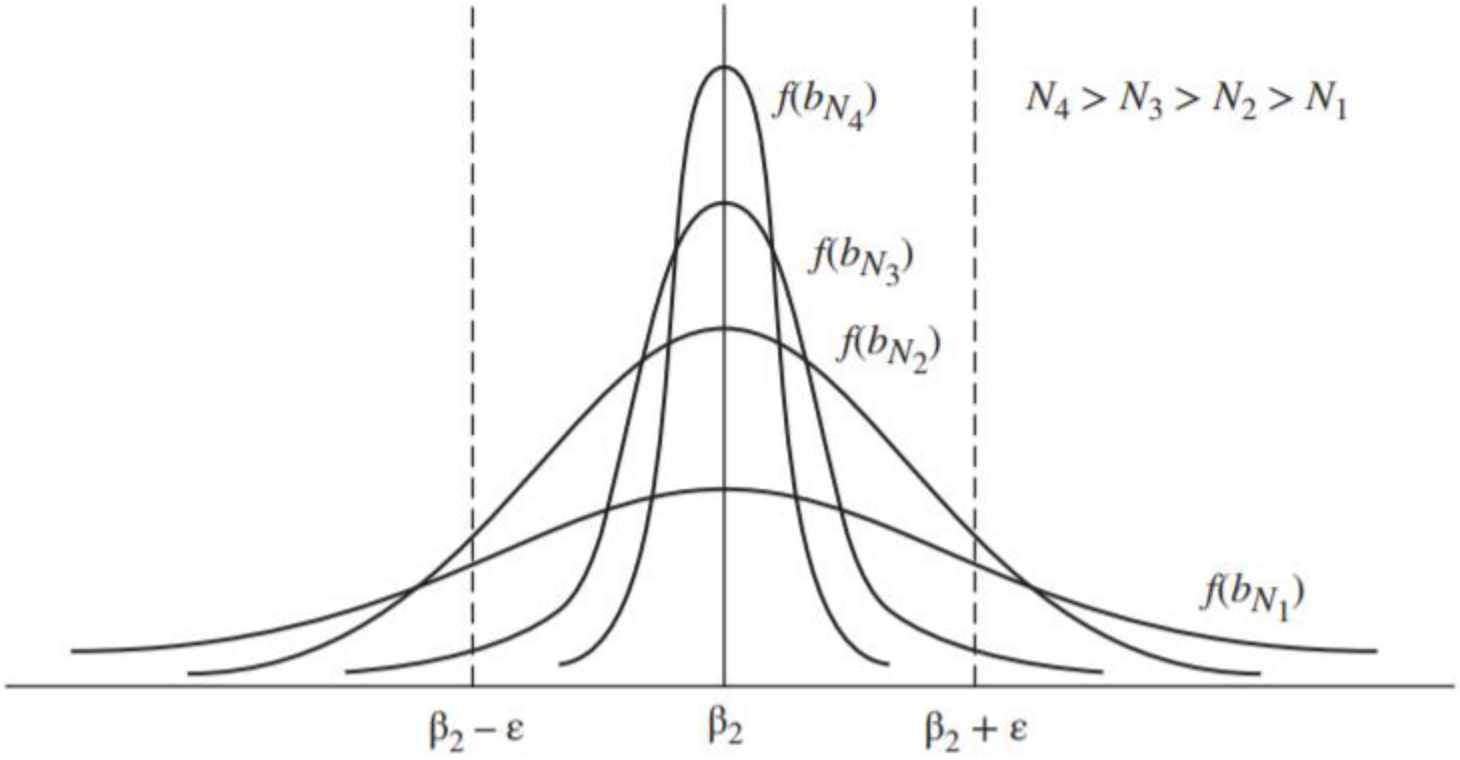


**FIGURE 5.5** An illustration of consistency.

# 5.7.2 Asymptotic Normality 1 of 3

- Large sample approximate distributions are called asymptotic distributions
- The need to use asymptotic distributions will become more urgent as we examine more complex models and estimators
- We consider instead the distribution of $\sqrt{N}b_2$
- It follows that $E(\sqrt{N}b_2) = \sqrt{N}\beta_2$
- And (5.35) $\sqrt{N}b_2 \sim (\sqrt{N}\beta_2, C_x)$

# 5.7.2 Asymptotic Normality 2 of 3

- Applying a central limit theorem to the sum $\sum_{i=1}^{N}(x_i - \bar{x})e_i/\sqrt{N}$ and using $[s_x^2]^{-1} \xrightarrow{P} C_x$ it can be shown that the statistic obtained by normalizing (5.35) so that it has mean zero and variance one, will be approximately normally distributed

- $\dfrac{\sqrt{N}(b_2 - \beta_2)}{\sqrt{\sigma^2 C_x}} \overset{a}{\sim} N(0,1)$

- Going one step further, there is an important theorem that says replacing unknown quantities with consistent estimators does not change the asymptotic distribution of a statistic

# 5.7.2 Asymptotic Normality 3 of 3

- Thus we can write (5.36) $t = \frac{\sqrt{N}(b_2 - \beta_2)}{\sqrt{\widehat{\sigma^2}/s_x^2}} \overset{a}{\sim} N(0,1)$
- This is precisely the t-statistic that we use for interval estimation and hypothesis testing. The result in (5.36) means that using it in large samples is justified when assumption MR6 is not satisfied
- This has been in terms of the distribution of $b_2$ in the simple regression model, the result also holds for estimators of the coefficients in the multiple regression model

# 5.7.3 Relaxing Assumptions 1 of 3

- **Weakening Strict Exogeneity: Cross-sectional Data**

- We now examine the implications of replacing $E(e_i|x_i) = 0$ with the weaker assumption: (5.38) E $(e_i)$ = 0 and cov($e_i$, $x_{ik}$) = 0 for i = 1, 2,…, N; k = 1, 2,…,K

- The seemingly innocuous weaker assumption in (5.38) means we can no longer show that the least squares estimator is unbiased

# 5.7.3 Relaxing Assumptions 2 of 3

- **Weakening Strict Exogeneity: Time-series Data**
- When we turn to time series data, the observations ($y_t$ , $x_t$) , t = 1, 2, …, T are not collected via random sampling and so it is no longer reasonable to assume they are independent
- The likely violation of cov($e_t$,$x_{st}$) = 0 for s ≠ t implies E ($e_t$|X ) = 0 will be violated, which in turn implies we cannot show that the least squares estimator is unbiased

# 5.7.3 Relaxing Assumptions 3 of 3

- To show consistency, we first assume the errors and the explanatory variables in the same time period are uncorrelated
- (5.44) $E(e_t) = 0$ and $\text{cov}(e_t, x_{tk}) = 0$ for t = 1, 2,…,N; k = 1, 2,…,K
- Errors and the explanatory variables that satisfy (5.44) are said to be contemporaneously uncorrelated
- We do not insist that $\text{cov}(e_t, x_{tk}) = 0$ for $t \neq s$

# 5.7.4 Inference for a Nonlinear Function of Coefficients

- The need for large sample or asymptotic distributions is not confined to situations where assumptions MR1–MR6 are relaxed
- Even if these assumptions hold, we still need to use large sample theory if a quantity of interest involves a nonlinear function of coefficients
- This problem in the next example

# Example 5.17 The Optimal Level of Advertising 1 of 2

- Advertising should be increased to the point where: $\beta_3 + 2\beta_4 ADVERT_0 = 1$

- A point estimate for $ADVERT_0$ is $\widehat{ADVERT}_0 = \frac{1-b_3}{2b_4} = \frac{1 - 12.1512}{2 \times (-2.76796)} = 2.014$

- For the estimated variance of the optimal level of advertising, we have

- $se(\hat{\lambda}) = \sqrt{0.016567} = 0.1287$

# Example 5.17 The Optimal Level of Advertising 2 of 2

- We are now in a position to get a 95% interval estimate for $\lambda = ADVERT_0$
- An approximate 95% interval estimate for $ADVERT_0$ is:
- $\left(\hat{\lambda} - t_{(0.975,71)} se(\hat{\lambda}), \hat{\lambda} + t_{(0.975,71)} se(\hat{\lambda})\right) = (2.014 - 1.994 \times 0.1287, 2.014 + 1.994 \times 0.1287) = (1.757, 2.271)$
- We estimate with 95% confidence that the optimal level of advertising lies between \$1757 and \$2271

# Key Words

- asymptotic normality
- BLU estimator
- consistency
- covariance matrix of least squares estimator
- critical value
- delta method
- error variance estimate
- error variance estimator
- explained sum of squares
- FWL theorem
- goodness-of-fit
- interaction variable
- interval estimate
- least squares estimates
- least squares estimation
- least squares estimators
- linear combinations
- marginal effect
- multiple regression model
- nonlinear functions
- one-tail test
- p-value
- polynomial
- regression coefficients
- standard errors
- sum of squared errors
- sum of squares due of regression
- testing significance
- total sum of squares
- two-tail test

# Copyright