Chapter 5
The Multiple Regression Model

Walter R. Paczkowski
Rutgers University

Chapter Contents

5.1
Introduction

- 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:

Eq. 5.1

$$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT$$

- 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

- β_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

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

- Similarly, β_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

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

- The econometric model is:

$$E(SALES) = \beta_1 + \beta_2 PRICE + \beta_3\ ADVERT$$

  - 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)*

Eq. 5.2 $$SALES = E\left(SALES\right) + e = \beta_1 + \beta_2 PRICE + \beta_3\ ADVERT + e$$

FIGURE 5.1 The multiple regression plane

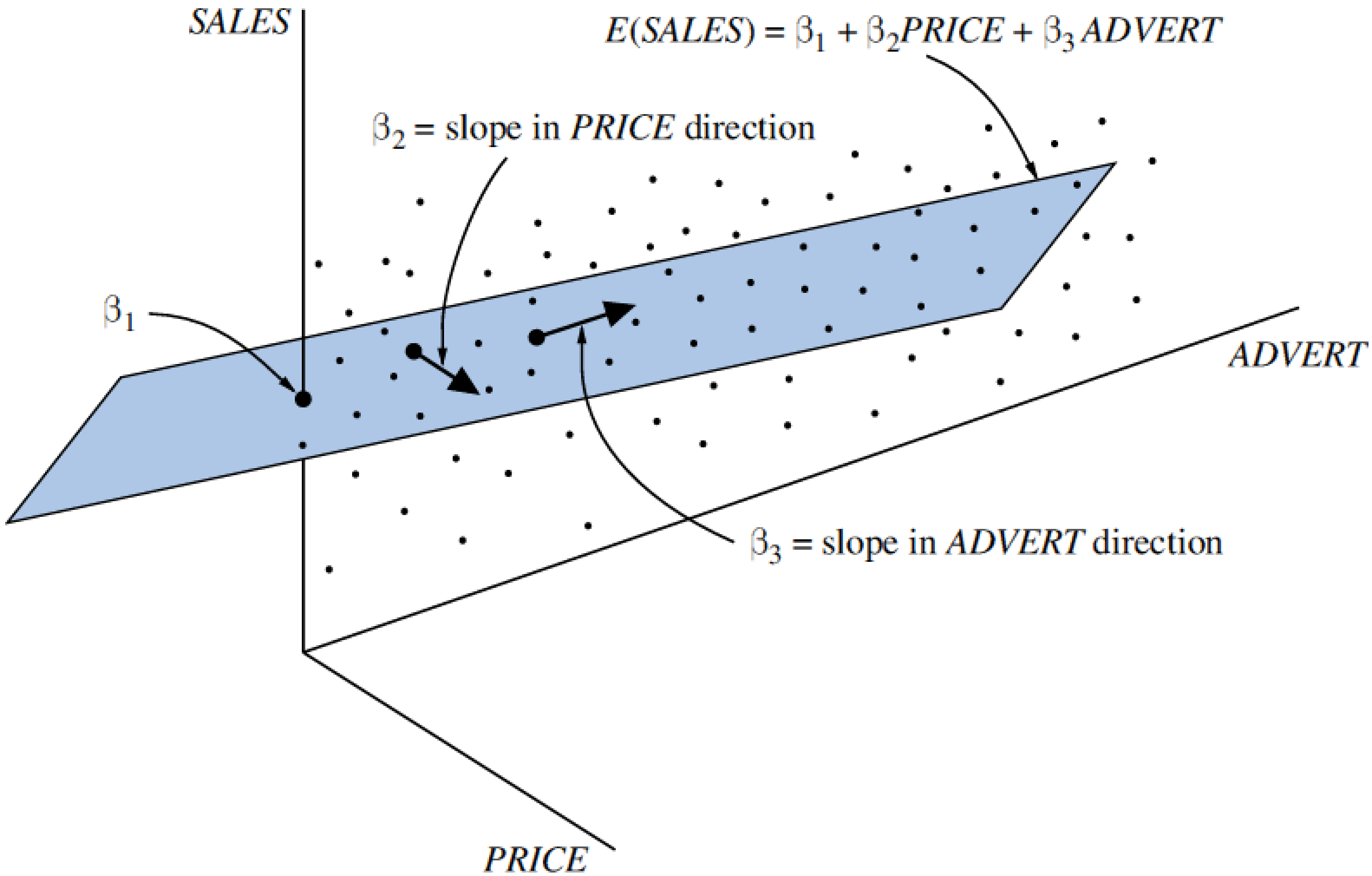


Table 5.1 Observations on Monthly Sales, Price, and Advertising in Big Andy's Burger Barn

City	*SALES* \$1,000 units	*PRICE* \$1 units	*ADVERT* \$1,000 units
1	73.2	5.69	1.3
2	71.8	6.49	2.9
3	62.4	5.63	0.8
4	67.4	6.22	0.7
5	89.3	5.02	1.5
.	.	.	.
.	.	.	.
.	.	.	.
73	75.4	5.71	0.7
74	81.3	5.45	2.0
75	75.0	6.05	2.2
Summary statistics			
Sample mean	77.37	5.69	1.84
Median	76.50	5.69	1.80
Maximum	91.20	6.49	3.10
Minimum	62.40	4.83	0.50
Std. Dev.	6.49	0.52	0.83

- In a general multiple regression model, a dependent variable *y* is related to a number of explanatory variables $x_2, x_3, \ldots, x_K$ through a linear equation that can be written as:

Eq. 5.3

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_K x_K + e$$

- A single parameter, call it β_k, measures the effect of a change in the variable x_k upon the expected value of y, all other variables held constant

$$\beta_k = \left.\frac{\Delta E(y)}{\Delta x_k}\right|_{\text{other xs held constant}} = \frac{\partial E(y)}{\partial x_k}$$

- The parameter β1 is the intercept term.
  - We can think of it as being attached to a variable x_1 that is always equal to 1
  - That is, $x_1 = 1$

- The equation for sales revenue can be viewed as a special case of Eq. 5.3 where $K = 3$, $y = SALES$, $x_1 = 1$, $x_2 = PRICE$ and $x_3 = ADVERT$

Eq. 5.4

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + e$$

- We make assumptions similar to those we made before:
  - $E(e) = 0$
  - $var(e) = \sigma^2$
    - Errors with this property are said to be **homoskedastic**
  - $cov(e_i, e_j) = 0$
  - $e \sim N(0, \sigma^2)$

- The statistical properties of y follow from those of e:
  - $E(y) = \beta_1 + \beta_2 x_2 + \beta_3 x_3$
  - $\text{var}(y) = \text{var}(e) = \sigma^2$
  - $\text{cov}(y_i, y_j) = \text{cov}(e_i, e_j) = 0$
  - $y \sim N[(\beta_1 + \beta_2 x_2 + \beta_3 x_3), \sigma^2]$
    - This is equivalent to assuming that $e \sim N(0, \sigma^2)$

- We make two assumptions about the explanatory variables:
  1. The explanatory variables are not random variables
     - We are assuming that the values of the explanatory variables are known to us prior to our observing the values of the dependent variable

- We make two assumptions about the explanatory variables (Continued):

  2. Any one of the explanatory variables is not an exact linear function of the others
     - This assumption is equivalent to assuming that no variable is redundant
     - If this assumption is violated – a condition called **exact collinearity** - the least squares procedure fails

ASSUMPTIONS of the Multiple Regression Model

MR1. $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + e_i, \ i = 1, \ldots, N$

MR2. $E(y_i) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} \Leftrightarrow E(e_i) = 0$

MR3. $\text{var}(y_i) = \text{var}(e_i) = \sigma^2$

MR4. $\text{cov}(y_i, y_j) = \text{cov}(e_i, e_j) = 0$

MR5. The values of each x_{tk} are not random and are not exact linear functions of the other explanatory variables

MR6. $y_i \sim N\left[(\beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}), \sigma^2\right] \Leftrightarrow e_i \sim N(0, \sigma^2)$

5.2

Estimating the Parameters of the Multiple Regression Model

- We will discuss estimation in the context of the model in Eq. 5.4, which we repeat here for convenience, with *i* denoting the *i*th observation

Eq. 5.4

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

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

- 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:

Eq. 5.5

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

- Formulas for b_1, b_2, and b_3, obtained by minimizing Eq. 5.5, 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

- Estimates along with their standard errors and the equation's R^2 are typically reported in equation format as:

Eq. 5.6

$$\widehat{SALES} = 118.91 - 7.908 PRICE + 1863 ADVERT \qquad R^2 = 0.448$$
$$(se) \qquad (6.35) \quad (1.096) \qquad (0.683)$$

Table 5.2 Least Squares Estimates for Sales Equation for Big Andy's Burger Barn

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	118.9136	6.3516	18.7217	0.0000
$PRICE$	−7.9079	1.0960	−7.2152	0.0000
$ADVERT$	1.8626	0.6832	2.7263	0.0080

$R^2 = 0.4483$ $SSE = 1718.943$ $\hat{\sigma} = 4.8861$ $s_y = 6.48854.$

- Interpretations of the results:
  1. The negative coefficient on *PRICE* suggests that demand is price elastic; we estimate that, with advertising held constant, an increase in price of $1 will lead to a fall in monthly revenue of $7,908
  2. The coefficient on advertising is positive; we estimate that with price held constant, an increase in advertising expenditure of $1,000 will lead to an increase in sales revenue of $1,863

- Interpretations of the results (Continued):

3. The estimated intercept implies that if both price and advertising expenditure were zero the sales revenue would be $118,914
   - Clearly, this outcome is not possible; a zero price implies zero sales revenue
   - In this model, as in many others, it is important to recognize that the model is an approximation to reality in the region for which we have data
   - Including an intercept improves this approximation even when it is not directly interpretable

- Using the model to predict sales if price is $5.50 and advertising expenditure is $1,200:

$$\begin{aligned} SALES &= 118.91 - 7.908 PRICE + 1.863 ADVERT \\ &= 118.914 - 7.9079 \times 5.5 + 1.8626 \times 1.2 \\ &= 77.656 \end{aligned}$$

  – The predicted value of sales revenue for *PRICE* = 5.5 and *ADVERT* =1.2 is $77,656.

- A word of caution is in order about interpreting regression results:
  - The negative sign attached to price implies that reducing the price will increase sales revenue.
  - If taken literally, why should we not keep reducing the price to zero?
  - Obviously that would not keep increasing total revenue
  - This makes the following important point:
    - Estimated regression models describe the relationship between the economic variables for values similar to those found in the sample data
    - Extrapolating the results to extreme values is generally not a good idea
    - Predicting the value of the dependent variable for values of the explanatory variables far from the sample values invites disaster

- We need to estimate the error variance, σ^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 σ^2 based on the squares of the least squares residuals:

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

- An estimator for σ^2 that uses the information from $\hat{e}_i^2$ and has good statistical properties is:

Eq. 5.7

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{e}_i^2}{N - K}$$

where K is the number of β parameters being estimated in the multiple regression model.

- For the hamburger chain example:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{75} \hat{e}_i^2}{N - K} = \frac{1718.943}{75 - 3} = 23.874$$

- Note that:

$$SSE = \sum_{i=1}^{N} \hat{e}_i^2 = 1718.943$$

  - Also, note that

$$\hat{\sigma} = \sqrt{23.874} = 4.8861$$

  - Both quantities typically appear in the output from your computer software
    - Different software refer to it in different ways.

5.3

Sampling Properties of the Least Squares Estimators

THE GAUSS-MARKOV THEOREM

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.

- If the errors are not normally distributed, then the least squares estimators are approximately normally distributed in large samples
  - What constitutes "large" is tricky
  - It depends on a number of factors specific to each application
  - Frequently, $N - K = 50$ will be large enough

- We can show that:

Eq. 5.8

$$\operatorname{var}(b_2) = \frac{\sigma^2}{(1 - r_{23}^2)\sum_{i=1}^{N}(x_{i2} - \bar{x}_2)^2}$$

where

Eq. 5.9

$$r_{23} = \frac{\sum(x_{i2} - \bar{x}_2)(x_{i3} - \bar{x}_3)}{\sqrt{\sum(x_{i2} - \bar{x}_2)^2 \sum(x_{i3} - \bar{x}_3)^2}}$$

- We can see that:
  1. Larger error variances σ^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

- We can arrange the variances and covariances in a matrix format:

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

- Using the hamburger data:

Eq. 5.10

$$\widehat{\text{cov}}(b_1, b_2, b_3) = \begin{bmatrix} 40.343 & -6.795 & -0.7484 \\ -6.795 & 1.201 & -0.0197 \\ -0.7484 & -0.0197 & 0.4668 \end{bmatrix}$$

- Therefore, we have:

$$\widehat{\text{var}}(b_1) = 40.343 \qquad \widehat{\text{cov}}(b_1, b_2) = -6.795$$

$$\widehat{\text{var}}(b_2) = 1.201 \qquad \widehat{\text{cov}}(b_1, b_3) = -0.7484$$

$$\widehat{\text{var}}(b_3) = 0.4668 \qquad \widehat{\text{cov}}(b_2, b_3) = -0.0197$$

- We are particularly interested in the standard errors:

$$se(b_1) = \sqrt{\widehat{\text{var}}(b_1)} = \sqrt{40.343} = 6.3516$$

$$se(b_2) = \sqrt{\widehat{\text{var}}(b_2)} = \sqrt{1.201} = 1.0960$$

$$se(b_3) = \sqrt{\widehat{\text{var}}(b_3)} = \sqrt{0.4668} = 0.6832$$

Table 5.3 Covariance Matrix for Coefficient Estimates

	C	*PRICE*	*ADVERT*
C	40.3433	−6.7951	−0.7484
PRICE	−6.7951	1.2012	−0.0197
ADVERT	−0.7484	−0.0197	0.4668

- 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:

$$y_i \sim N\left[(\beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}), \sigma^2\right] \Leftrightarrow e_i \sim N(0, \sigma^2)$$

- Since the least squares estimators are linear functions of dependent variables, it follows that the least squares estimators are also normally distributed:

$$b_k \sim N\left[\beta_k, \operatorname{var}\left(b_k\right)\right]$$

- We can now form the standard normal variable *Z*:

Eq. 5.11

$$z = \frac{b_k - \beta_k}{\sqrt{\text{var}\left(b_k\right)}} \sim N\left(0,1\right), \text{ for } k = 1, 2, \ldots, K$$

- Replacing the variance of b_k with its estimate:

Eq. 5.12

$$t = \frac{b_k - \beta_k}{\sqrt{\widehat{\text{var}}(b_k)}} = \frac{b_k - \beta_k}{\text{se}(b_k)} \sim t_{(N-K)}$$

  – Notice that the number of degrees of freedom for *t*-statistics is *N* - *K*

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

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

- And then we have

Eq. 5.13

$$t = \frac{\hat{\lambda} - \lambda}{se\left(\hat{\lambda}\right)} = \frac{\sum c_k b_k - \sum c_k \beta_k}{\text{se}\left(\sum c_k b_k\right)} \sim t_{(N-K)}$$

- If = 3, the we have:

$$\operatorname{se}\left(c_1 b_1 + c_2 b_2 + c_3 b_K\right) = \sqrt{\widehat{\operatorname{var}}\left(c_1 b_1 + c_2 b_2 + c_3 b_K\right)}$$

where

Eq. 5.14

$$\widehat{\operatorname{var}}\left(c_1 b_1 + c_2 b_2 + c_3 b_K\right) = c_1^2 \widehat{\operatorname{var}}\left(b_1\right) + c_2^2 \widehat{\operatorname{var}}\left(b_2\right) + c_3^2 \widehat{\operatorname{var}}\left(b_3\right)$$

$$+2c_1c_2 \widehat{\operatorname{cov}}\left(b_1, b_2\right) + 2c_1c_3 \widehat{\operatorname{cov}}\left(b_1, b_3\right) + 2c_2c_3 \widehat{\operatorname{cov}}\left(b_2, b_3\right)$$

- What happens if the errors are not normally distributed?
  - Then the least squares estimator will not be normally distributed and Eq. 5.11, Eq. 5.12, and Eq. 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 Eq. 5.12 and Eq. 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
Interval Estimation

- We now examine how we can do interval estimation and hypothesis testing

- For the hamburger example, we need:

Eq. 5.15

$$P(-t_c < t_{(72)} < t_c) = .95$$

- Using t_c = 1.993, we can rewrite (5.15) as:

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

- Rearranging, we get:

$$P\left[b_2 - 1.993 \times \text{se}(b_2) \;\leq\; \beta_2 \;\leq\; b_2 + 1.993 \times \text{se}(b_2)\right] = .95$$

  – Or, just writing the end-points for a 95% interval:

Eq. 5.16

$$\left[b_2 - 1.993 \times \text{se}(b_2),\; b_2 + 1.993 \times \text{se}(b_2)\right]$$

- Using our data, we have b_2 = -7.908 and $se(b_2)$ = 1.096, so that:

$$\left(-7.9079-1.993\times1.096, -7.9079+1.993\times1.096\right)=\left(10.093,-5.723\right)$$

  - This interval estimate suggests that decreasing price by $1 will lead to an increase in revenue somewhere between $5,723 and $10,093.
    - In terms of a price change whose magnitude is more realistic, a 10-cent price reduction will lead to a revenue increase between $572 and $1,009

- Similarly for advertising, we get:

$$\left(1.8626 - 1.9935 \times 0.6832, 1.8626 + 1.9935 \times 0.6832\right) = \left(0.501, 3.225\right)$$

  - We estimate that an increase in advertising expenditure of $1,000 leads to an increase in sales revenue of between $501 and $3,225
  - This interval is a relatively wide one; it implies that extra advertising expenditure could be unprofitable (the revenue increase is less than $1,000) or could lead to a revenue increase more than three times the cost of the advertising

- We write the general expression for a 100(1-α)% confidence interval as:

$$\left(b_k - t_{(1-\alpha/2,\, N-K)} \times \text{se}(b_k),\ \ b_k + t_{(1-\alpha/2,\, N-K)} \times \text{se}(b_k)\right)$$

- Suppose Big Andy wants to increase advertising expenditure by $800 and drop the price by 40 cents.
  - Then the change in expected sales is:

$$\lambda = E\left(SALES_1\right) - E\left(SALES_0\right)$$
$$= \left[\beta_1 + \beta_2\left(PRICE_0 - 0.4\right) + \beta_3\left(ADVERT_0 + 0.8\right)\right]$$
$$- \left[\beta_1 + \beta_2 PRICE_0 + \beta_3 ADVERT_0\right]$$
$$= -0.4\beta_2 + 0.8\beta_3$$

- 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:

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

- The standard error is:

$$\begin{aligned}\text{se}\left(-0.4\text{b}_2+0.8\text{b}_3\right) &= \sqrt{\text{var}\left(-0.4\text{b}_2+0.8\text{b}_3\right)} \\ &= \sqrt{\left(-0.4\right)^2 \widehat{\text{var}}\left(b_2\right)+\left(0.8\right)^2 \widehat{\text{var}}\left(b_3\right)-2\times-0.4\times0.8\times\widehat{\text{cov}}\left(b_2,b_3\right)} \\ &= \sqrt{0.16\times1.2012+0.64\times0.4668-0.64\times\left(-0.0197\right)} \\ &= 0.7096\end{aligned}$$

- The 90% interval is then:

$$\left(4.6532-1.666\times0.7096, 4.6532+1.666\times0.7096\right)=\left(3.471, 5.835\right)$$

  – We estimate, with 90% confidence, that the expected increase in sales will lie between \$3,471 and \$5,835

5.5
Hypothesis Testing

COMPONENTS OF HYPOTHESIS TESTS

1. A null hypothesis H_0
2. An alternative hypothesis H_1
3. A test statistic
4. A rejection region
5. A conclusion

- We need to ask whether the data provide any evidence to suggest that y is related to each of the explanatory variables
  - 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

- Null hypothesis:

$$H_0 : \beta_k = 0$$

- Alternative hypothesis:

$$H_1 : \beta_k \neq 0$$

- Test statistic:

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

- t values for a test with level of significance α:

$$t_c = t_{(1-\alpha/2,\, N-K)} \text{ and } -t_c = t_{(\alpha/2,\, N-K)}$$

- For our hamburger example, we can conduct a test that sales revenue is related to price:
  1. The null and alternative hypotheses are:
  $$H_0: \beta_2 = 0 \text{ and } H_1: \beta_2 \neq 0$$
  2. The test statistic, if the null hypothesis is true, is:
  $$t = b_2 / \text{se}(b_2) \sim t_{(N-K)}$$
  3. Using a 5% significance level (α=.05), and 72 degrees of freedom, the critical values that lead to a probability of 0.025 in each tail of the distribution are:
  $$t_{(.975,72)} = 1.993 \text{ and } t_{(.025,72)} = -1.993$$

- For our hamburger example (Continued) :
  4. The computed value of the *t*-statistic is:

$$t = \frac{-7.908}{1.096} = -7.215$$

and the *p*-value from software is:

$$P\left(t_{(72)} > 7.215\right) + P\left(t_{(72)} < -7.215\right) = 2 \times (2.2 \times 10^{-10}) = 0.000$$

  5. Since -7:215 < -1.993, we reject H_0: $\beta_2 = 0$ and conclude that there is evidence from the data to suggest sales revenue depends on price
     - Using the *p*-value to perform the test, we reject H_0 because 0.000 < 0.05 .

- Similarly, we can conduct a test that sales revenue is related to advertising expenditure:
  1. The null and alternative hypotheses are:
  $$H_0 : \beta_3 = 0 \text{ and } H_1 : \beta_3 \neq 0$$
  2. The test statistic, if the null hypothesis is true, is:
  $$t = b_3 / \text{se}(b_3) \sim t_{(N-K)}$$
  3. Using a 5% significance level (α=.05), and 72 degrees of freedom, the critical values that lead to a probability of 0.025 in each tail of the distribution are:
  $$t_{(.975,72)} = 1.993 \text{ and } t_{(.025,72)} = -1.993$$

- For our hamburger example (Continued):

4. The computed value of the *t*-statistic is:

$$t = \frac{1.8626}{0.6832} = 2.726$$

and the *p*-value from software is:

$$P\left(t_{(72)} > 2.726\right) + P\left(t_{(72)} < -2.726\right) = 2 \times 0.004 = 0.008$$

5. Since 2.726 > 1.993, we reject H_0: $\beta_3 = 0$: the data support the conjecture that revenue is related to advertising expenditure

   - Using the *p*-value to perform the test, we reject H_0 because $0.008 < 0.05$.

- We now are in a position to state the following questions as testable hypotheses and ask whether the hypotheses are compatible with the data
  1. Is demand price-elastic or price-inelastic?
  2. Would additional sales revenue from additional advertising expenditure cover the costs of the advertising?

- For the demand elasticity, we wish to know if:
  – $\beta_2 \geq 0$: a decrease in price leads to a decrease in sales revenue (demand is price-inelastic or has an elasticity of unity), or
  – $\beta_2 < 0$: a decrease in price leads to a decrease in sales revenue (demand is price-inelastic)

- As before:
  1. The null and alternative hypotheses are:

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

     $H_1 : \beta_2 < 0$ (demand is elastic)

  2. The test statistic, if the null hypothesis is true, is:

     $$t = b_2 / \text{se}(b_2) \sim t_{(N-K)}$$

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

- Hypothesis test (Continued):

4. The test statistic is:

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

and the *p*-value is:

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

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

- The other hypothesis of interest is whether an increase in advertising expenditure will bring an increase in sales revenue that is sufficient to cover the increased cost of advertising
  - Such an increase will be achieved if $\beta_3 > 1$

- As before:
  1. The null and alternative hypotheses are:

$$H_0 : \beta_3 \leq 1$$

$$H_1 : \beta_3 > 1$$

  2. The test statistic, if the null hypothesis is true, is:

$$t = \frac{b_3 - 1}{\text{se}(b_3)} \sim t_{(N-K)}$$

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

- Hypothesis test (Continued):

4. The test statistic is:

$$t = \frac{b_3 - \beta_3}{se\left(b_2\right)} = \frac{1.8626 - 1}{0.6832} = 1.263$$

and the *p*-value is:

$$P\left(t_{(72)} > 1.263\right) = 0.105$$

5. Since 1.263 < 1.666, we do not reject H_0

- The marketing adviser claims that dropping the price by 20 cents will be more effective for increasing sales revenue than increasing advertising expenditure by $500
  - In other words, she claims that $-0.2\beta_2 > 0.5\beta_3$, or $-0.2\beta_2 - 0.5\beta_3 > 0$
  - We want to test a hypothesis about the linear combination $-0.2\beta_2 - 0.5\beta_3$

- As before:
  1. The null and alternative hypotheses are:

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

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

  2. The test statistic, if the null hypothesis is true, is:

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

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

- We need the standard error:

$$\text{se}(-0.2b_2 - 0.5b_3) = \sqrt{\widehat{\text{var}}\left(\text{se}(-0.2b_2 - 0.5b_3)\right)}$$

$$= \sqrt{(-0.2)^2 \widehat{\text{var}}(b_2) + (-0.5)^2 \widehat{\text{var}}(b_3) + 2 \times (-0.2) \times (-0.5) \widehat{\text{cov}}(b2, b3)}$$

$$= \sqrt{0.04 \times 1.2012 + 0.25 \times 0.4668 + 0.2 \times (-0.0197)}$$

$$= 0.4010$$

- Hypothesis test (Continued):

4. The test statistic is:

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

and the *p*-value is:

$$P\left(t_{(72)} > 1.622\right) = 0.055$$

5. Since 1.622 < 1.666, we do not reject H_0

5.6

Polynomial Equations

- We have studied the multiple regression model

Eq. 5.17

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + K + \beta_K x_K + e$$

– 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

$$y = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3 + e.$$

- Consider the average cost equation

Eq. 5.18

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

  – And the total cost function

Eq. 5.19

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

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

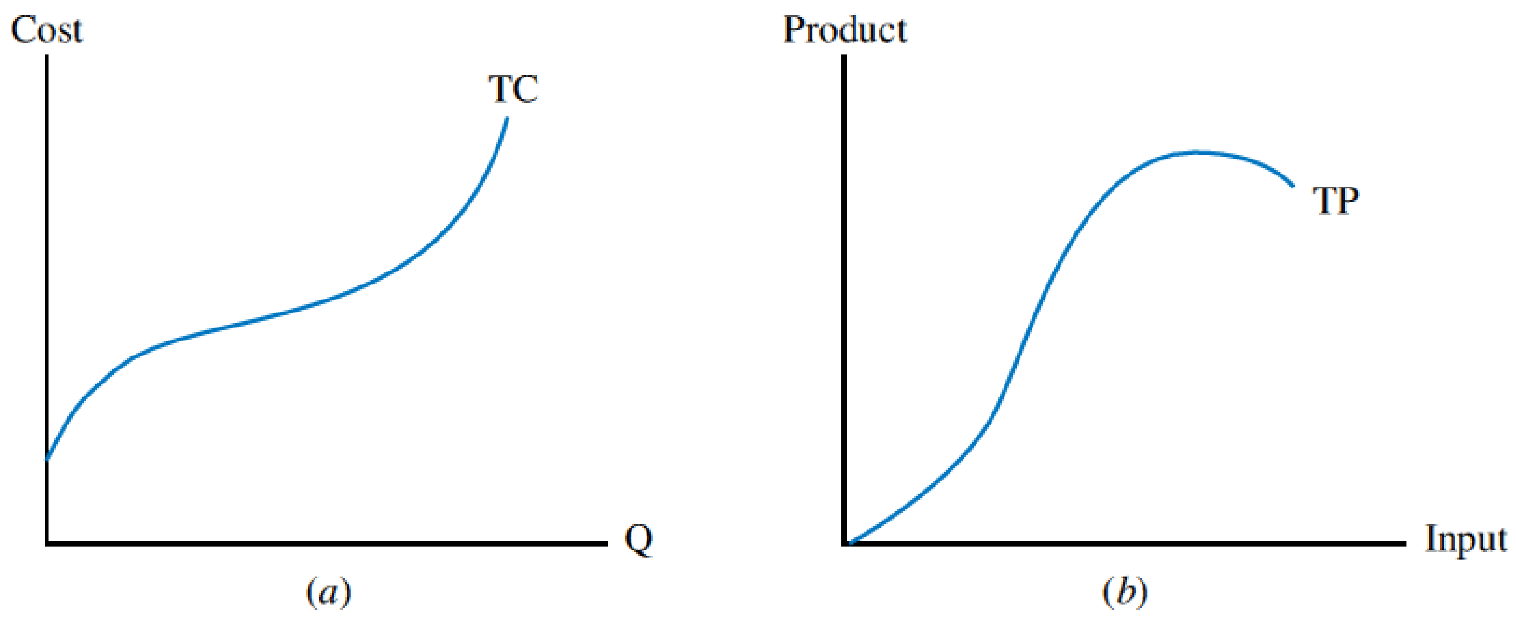
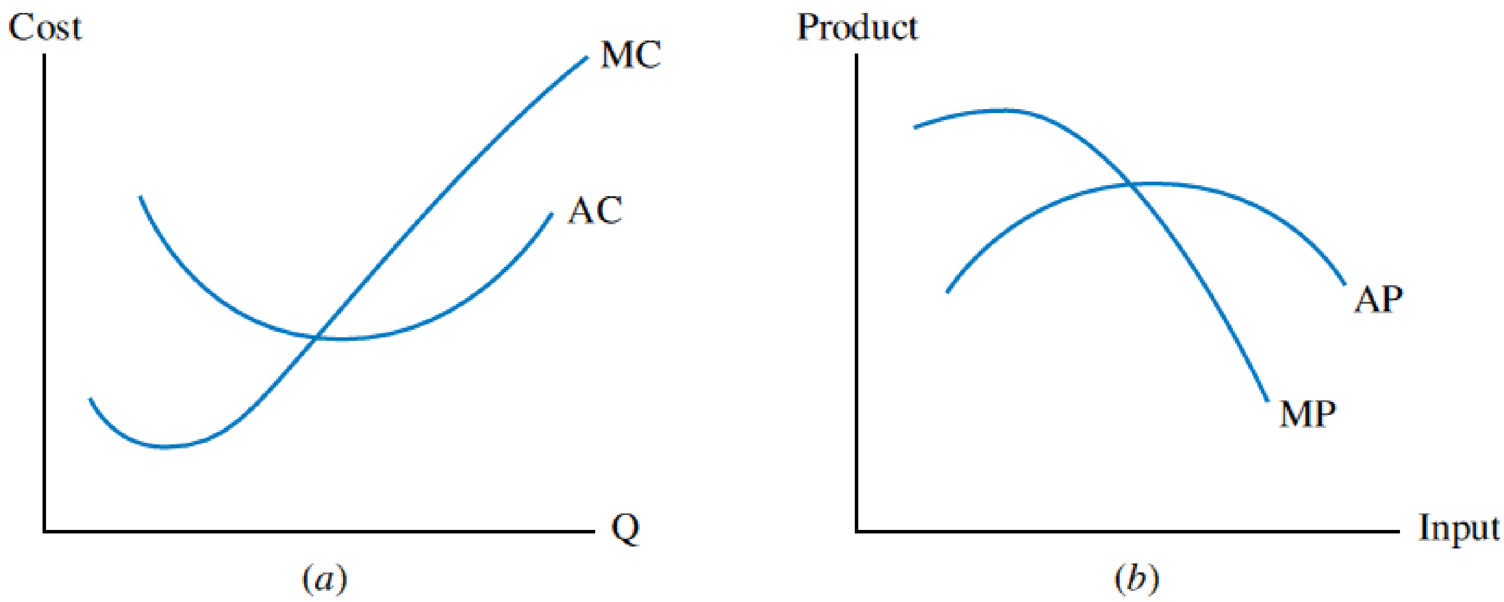


FIGURE 5.3 Average and marginal (a) cost curves and (b) product curves



- 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

Eq. 5.20

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

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

Eq. 5.21

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

- The slope of the average cost curve Eq. 5.18 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 $\alpha_4 > 0$

- The slope of the total cost curve Eq. 5.19 ,which is the marginal cost, is:

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

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

- It is sometimes true that having a variable and its square or cube in the same model causes **collinearity** problems
  - This will be discussed in Chapter 6

- The linear sales model with the constant slope β_3 for advertising

$$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT + e$$

does not capture diminishing returns in advertising expenditure

- A new, better model might be:

Eq. 5.22

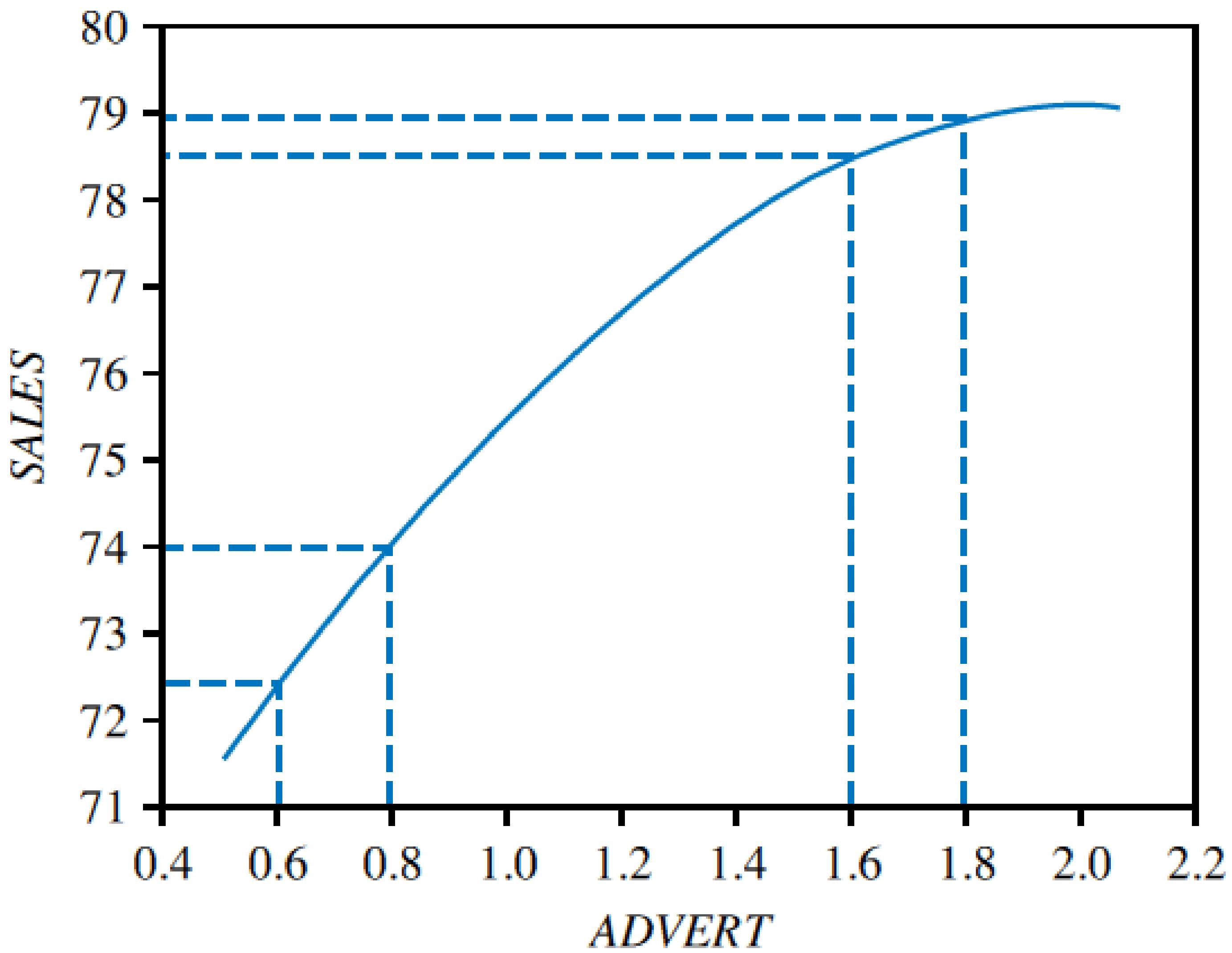
$$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT + \beta_4 ADVERT^2 + e$$

- The change in expected sales to a change in advertising is:

Eq. 5.23

$$\left.\frac{\Delta E(SALES)}{\Delta ADVERT}\right|_{(PRICE \text{ held constant})} = \frac{\partial E(SALES)}{\partial ADVERT} = \beta_3 + 2\beta_4 ADVERT$$

FIGURE 5.4 A model where sales exhibits diminishing returns to advertising expenditure



- We refer to

$$\partial E(SALES)/\partial ADVERT$$

as the marginal effect of advertising on sales

- The least squares estimates are:

Eq. 5.24

$$\widehat{SALES} = 109.72 - 7.640PRICE + 12.151ADVERT - 2.768ADVERT^2$$
$$(\text{se}) \qquad (6.80) \qquad (1.046) \qquad (3.556) \qquad (0.941)$$

- The estimated response of sales to advertising is:

$$\frac{\partial \widehat{SALES}}{\partial ADVERT} = 12.151 - 5.536 ADVERT$$

- Substituting, we find that when advertising is at its minimum value in the sample of \$500 (*ADVERT* = 0.5), the marginal effect of advertising on sales is 9.383
  - When advertising is at a level of \$2,000 (*ADVERT* = 2), the marginal effect is 1.079
  - Allowing for diminishing returns to advertising expenditure has improved our model both statistically and in terms of meeting expectations about how sales will respond to changes in advertising

- The marginal benefit from advertising is the marginal revenue from more advertising
  - The required marginal revenue is given by the marginal effect of more advertising
    $\beta_3 + 2\beta_4 ADVERT$
  - The marginal cost of \$1 of advertising is \$1 plus the cost of preparing the additional products sold due to effective advertising
  - Ignoring the latter costs, the marginal cost of \$1 of advertising expenditure is \$1

- Advertising should be increased to the point where

$$\beta_3 + 2\beta_4 ADVERT_0 = 1$$

with $ADVERT_0$ denoting the optimal level of advertising

- Using the least squares estimates, a point estimate of $ADVERT_0$ is:

$$\widehat{ADVERT}_0 = \frac{1 - b_3}{2b_4} = \frac{1 - 12.1512}{2 \times (-2.76796)} = 2.014$$

implying that the optimal monthly advertising expenditure is $2,014

- Variances of nonlinear functions are hard to derive
  - Recall that the variance of a linear function, say, $c_3b_3 + c_4b_4$, is:

Eq. 5.25

$$\operatorname{var}\left(c_3b_3 + c_4b_4\right) = c_3^2 \operatorname{var}\left(b_3\right) + c_4^2 \operatorname{var}\left(b_4\right) + 2c_3c_4 \operatorname{cov}\left(b_3, b_4\right)$$

- Suppose $\lambda = (1 - \beta_3)/2\beta_4$ and

$$\hat{\lambda} = \frac{1 - b_3}{2b_4}$$

  - Then, the approximate variance expression is:

Eq. 5.26

$$\operatorname{var}\left(\hat{\lambda}\right) = \left(\frac{\partial\lambda}{\partial\beta_3}\right)^2 \operatorname{var}(b_3) + \left(\frac{\partial\lambda}{\partial\beta_4}\right)^2 \operatorname{var}(b_4) + 2\left(\frac{\partial\lambda}{\partial\beta_3}\right)\left(\frac{\partial\lambda}{\partial\beta_4}\right)\operatorname{cov}(b_3, b_4)$$

  - Using Eq. 5.26 to find an approximate expression for a variance is called the **delta method**

- The required derivatives are:

$$\frac{\partial \lambda}{\partial \beta_3} = -\frac{1}{2\beta_4}, \quad \frac{\partial \lambda}{\partial \beta_4} = -\frac{1-\beta_3}{2\beta_4^2}$$

- Thus, for the estimated variance of the optimal level of advertising, we have:

$$\widehat{\text{var}}\left(\hat{\lambda}\right)=\left(-\frac{1}{2b_4}\right)^2\widehat{\text{var}}\left(b_3\right)+\left(-\frac{1-b_3}{2b_4^2}\right)^2\widehat{\text{var}}\left(b_4\right)+2\left(-\frac{1}{2b_4}\right)\left(-\frac{1-b_3}{2b_4^2}\right)\widehat{\text{cov}}\left(b_3,b_4\right)$$

$$=\left(\frac{1}{2\times 2.768}\right)^2\times 12.646+\left(\frac{1-12.151}{2\times 2.768^2}\right)^2\times 0.88477$$

$$+2\left(\frac{1}{2\times 2.768}\right)\left(\frac{1-12.151}{2\times 2.768^2}\right)\times 3.2887$$

$$=0.016567$$

and

$$\text{se}\left(\hat{\lambda}\right)=\sqrt{0.016567}=0.1287$$

- An approximate 95% interval estimate for $ADVERT_0$ is:

$$\left(\hat{\lambda} - t_{(0.975,71)}\text{se}\left(\hat{\lambda}\right), \hat{\lambda} + t_{(0.975,71)}\text{se}\left(\hat{\lambda}\right)\right)$$

$$= \left(2.014 - 1.994 \times 0.1287, 2.014 + 1.994 \times 0.1287\right)$$

$$= \left(1.757,\ 2.271\right)$$

  - We estimate with 95% confidence that the optimal level of advertising lies between \$1,757 and \$2,271

5.7
Interaction Variables

- Suppose that we wish to study the effect of income and age on an individual's expenditure on pizza
  - An initial model would be:

Eq. 5.27

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + e$$

- Implications of this model are:
  1. $\partial E(PIZZA)/\partial AGE = \beta_2$: For a given level of income, the expected expenditure on pizza changes by the amount β_2 with an additional year of age
  2. $\partial E(PIZZA)/\partial INCOME = \beta_3$: For individuals of a given age, an increase in income of $1,000 increases expected expenditures on pizza by β_3

Table 5.4 Pizza Expenditure Data

PIZZA	*INCOME*	*AGE*
109	19.5	25
0	39.0	45
0	15.6	20
108	26.0	28
220	19.5	25

- The estimated model is:

$$PIZZA = 342.88 - 7.576AGE + 1.832INCOME$$
$$\text{(t)} \qquad\qquad (-3.27) \qquad (3.95)$$

  - The signs of the estimated parameters are as we anticipated
    - Both *AGE* and *INCOME* have significant coefficients, based on their *t*-statistics

- It is not reasonable to expect that, regardless of the age of the individual, an increase in income by $1,000 should lead to an increase in pizza expenditure by $1.83?
  - It would seem more reasonable to assume that as a person grows older, his or her 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

- We will add the interaction variable (*AGE* x *INCOME*) to the regression model
  - The new model is:

Eq. 5.28

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + \beta_4 \left(AGE \times INCOME \right) + e$$

- Implications of this revised model are:
  1. $\partial E(PIZZA)/\partial AGE = \beta_2 + \beta_4 INCOME$
  2. $\partial E(PIZZA)/\partial INCOME = \beta_3 + \beta_4 AGE$

- The estimated model is:

$$PIZZA = 161.47 - 2.977\,AGE + 6.980\,INCOME - 0.1232\left(AGE \times INCOME\right)$$
$$(t) \qquad \left(\text{-}0.89\right) \qquad \left(2.47\right) \qquad \left(-1.85\right)$$

- The estimated marginal effect of age upon pizza expenditure for two individuals—one with \$25,000 income and one with \$90,000 income is:

$$\frac{\partial \hat{E}(PIZZA)}{\partial AGE} = b_2 + b_4 INCOME$$

$$= -2.977 - 0.1232 INCOME$$

$$= \begin{cases} -6.06 & \text{for } INCOME = 25 \\ -14.07 & \text{for } INCOME = 90 \end{cases}$$

  – We expect that an individual with \$25,000 income will reduce pizza expenditures by \$6.06 per year, whereas the individual with \$90,000 income will reduce pizza expenditures by \$14.07 per year

- Consider a wage equation where ln(*WAGE*) depends on years of education (*EDUC*) and years of experience (*EXPER*):

Eq. 5.29

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + e$$

  – If we believe the effect of an extra year of experience on wages will depend on the level of education, then we can add an interaction variable

Eq. 5.30

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 (EDUC \times EXPER) + e$$

- The effect of another year of experience, holding education constant, is roughly:

$$\left.\frac{\Delta \ln(WAGE)}{\Delta EXPER}\right|_{EDUC \text{ fixed}} = \beta_3 + \beta_4 EDUC$$

  – The approximate percentage change in wage given a one-year increase in experience is $100(\beta_3+\beta_4 EDUC)\%$

- An estimated model is:

$$\widehat{\ln(WAGE)} = 1.392 + 0.09494 EDUC + 0.00633 EXPER - 0.0000364 (EDUC \times EXPER)$$

- If there is an interaction and a quadratic term on the right-hand side, as in

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \beta_4 (EDUC \times EXPER) + \beta_5 EXPER^2 + e$$

then we find that a one-year increase in experience leads to an approximate percentage wage change of:

$$\%\Delta WAGE \cong 100(\beta_3 + \beta_4 EDUC + 2\beta_5 EXPER)\%$$

5.8
Measuring Goodness-of-fit

- In the multiple regression model the R^2 is relevant and the same formulas are valid, but now we talk of the proportion of variation in the dependent variable explained by all the explanatory variables included in the linear model

- The coefficient of determination is:

Eq. 5.31

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

$$= 1 - \frac{SSE}{SST}$$

$$= 1 - \frac{\sum_{i=1}^{N} \hat{e}_i^{\,2}}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

- The predicted value of *y* is:

$$\hat{y}_i = b_1 + b_2 x_{i2} + b_3 x_{i3} + \cdots + b_K x_{iK}$$

– Recall that:

$$s_y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2} = \sqrt{\frac{SST}{N-1}}$$

– Then:

$$SST = \left(N-1\right)s_y^2$$

- For the hamburger example:.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \hat{e}_i^{\,2}}{\sum_{i=1}^{N} \left(y_i - \bar{y}\right)^2} 1 - \frac{1718.943}{3115.482} = 0.448$$

- Interpretation
  - 44.8% of the variation in sales revenue is explained by the variation in price and by the variation in the level of advertising expenditure
  - In our sample, 55.2% of the variation in revenue is left unexplained and is due to variation in the error term or to variation in other variables that implicitly form part of the error term
  - Adding the square of advertising to the Burger Barn model (see Eq. 5.24) increased the R^2 to 0.508
    - Thus an additional 6% of the variation in sales is explained by including this variable

- If the model does not contain an intercept parameter, then the measure R^2 given in Eq. 5.31 is no longer appropriate
  - The reason it is no longer appropriate is that without an intercept term in the model,

$$\sum_{i=1}^{N}(y_i - \bar{y})^2 \neq \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{N}\hat{e}_i^{\,2}$$

so $SST \neq SSR + SSE$

- Under these circumstances it does not make sense to talk of the proportion of total variation that is explained by the regression
  - When your model does not contain a constant, it is better not to report R^2
    - Even if your computer displays one

Key Words

Keywords

- BLU estimator
- covariance matrix of least squares estimator
- critical value
- delta method
- error variance estimate
- error variance estimator
- goodness-of-fit
- interaction variable
- interval estimate
- least squares estimates
- least squares estimation
- least squares estimators
- linear combinations
- marginal effects
- multiple regression model
- Nonlinear functions
- one-tailed test
- *p*-value
- Polynomial
- regression coefficients
- standard errors
- sum of squared errors
- sum of squares of regression
- testing significance
- total sum of squares
- two-tailed test

Appendices

- 5A Derivation of Least Squares Estimators
- 5B Large Sample Analysis

- Following a procedure similar to the single variable model, we have to minimize:

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

  - To get:

$$\frac{\partial S}{\partial \beta_1}=2N\beta_1+2\beta_2\sum x_{i2}+2\beta_3\sum x_{i3}-2\sum y_i$$

$$\frac{\partial S}{\partial \beta_2}=2\beta_1\sum x_{i2}+2\beta_2\sum x_{i2}^2+2\beta_3\sum x_{i2}x_{i3}-2\sum x_{i2}y_i$$

$$\frac{\partial S}{\partial \beta_3}=2\beta_1\sum x_{i3}+2\beta_2\sum x_{i2}x_{i3}+2\beta_3\sum x_{i3}^2-2\sum x_{i3}y_i$$

- We now can get three simultaneous **normal equations**:

Eq. 5A.1

$$Nb_1 + \sum x_{i2} b_2 + \sum x_{i3} b_3 = \sum y_i$$

$$\sum x_{i2} b_1 + \sum x_{i2}^2 b_2 + \sum x_{i2} x_{i3} b_3 = \sum x_{i2} y_i$$

$$\sum x_{i3} b_1 + \sum x_{i2} x_{i3} b_2 + \sum x_{i3}^2 b_3 = \sum x_{i3} y_i$$

- To help solve them, let:

$$y_i^* = y_i - \bar{y}, \quad x_{i2}^* = x_{i2} - \bar{x}_2, \quad x_{i3}^* = x_{i3} - \bar{x}_3$$

  - Then the least squares estimators are:

$$b_1 = \bar{y} - b_2\bar{x}_2 - b_3\bar{x}_3$$

$$b_2 = \frac{\left(\sum y_i^* x_{i2}^*\right)\left(\sum x_{i3}^{*2}\right) - \left(\sum y_i^* x_{i3}^*\right)\left(\sum x_{i2}^* x_{i3}^*\right)}{\left(\sum x_{i2}^{*2}\right)\left(\sum x_{i3}^{*2}\right) - \left(\sum x_{i2}^* x_{i3}^*\right)^2}$$

$$b_3 = \frac{\left(\sum y_i^* x_{i3}^*\right)\left(\sum x_{i2}^{*2}\right) - \left(\sum y_i^* x_{i2}^*\right)\left(\sum x_{i3}^* x_{i2}^*\right)}{\left(\sum x_{i2}^{*2}\right)\left(\sum x_{i3}^{*2}\right) - \left(\sum x_{i2}^* x_{i3}^*\right)^2}$$

- In the multiple regression model, if assumptions MR1–MR5 hold (or SR1–SR5 in the simple regression model) we are able to show that the least squares estimators are Best, Linear, Unbiased Estimators (*BLUE*)
  – These properties are called "finite sample" properties because they do not depend on the sample size N, and will hold if the sample is any size $N > K$
- There are additional properties of the least squares estimator that can be established if samples are imagined becoming infinitely large
  – In econometrics and statistics these are called **asymptotic properties**, with the term asymptotic implying the analysis of limiting behavior, here as $N \rightarrow \infty$

- The probability of obtaining an estimate "close" to β_2 is:

Eq. 5B.1

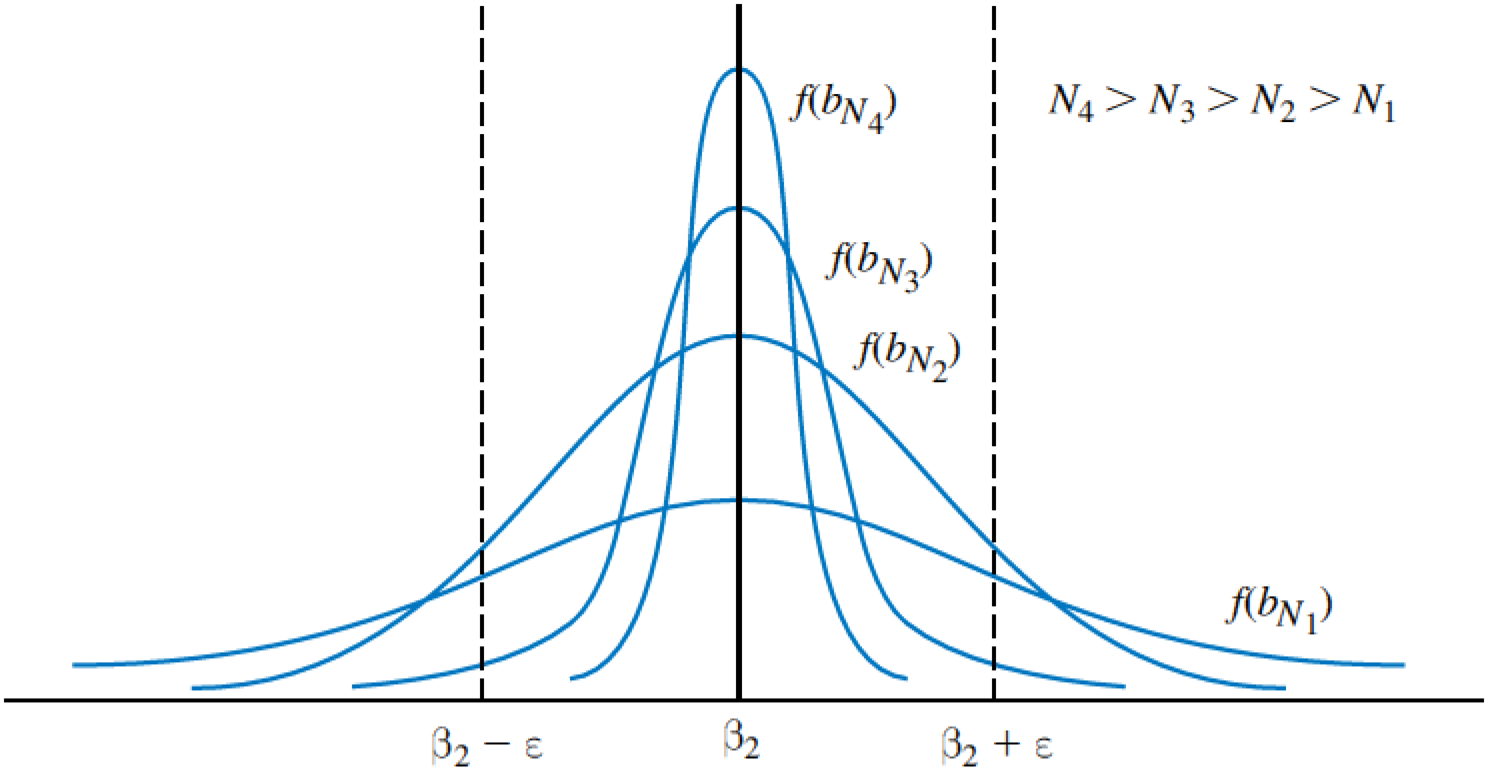
$$P\left(\beta_2 - \varepsilon \le b_2 \le \beta_2 + \varepsilon\right)$$

- An estimator is said to be **consistent** if this probability converges to 1 as the sample size $N \rightarrow \infty$

Eq. 5B.2

$$\lim_{N\rightarrow\infty} P\left(\beta_2 - \varepsilon \le b_2 \le \beta_2 + \varepsilon\right) = 1$$

FIGURE 5B.1 An illustration of consistency



- The property of consistency applies to many estimators, even ones that are biased in finite samples
  - Example: the estimator $\hat{\beta}_2 = b_2 + 1/N$ is a biased estimator
    - The bias is

$$\text{bias}\left(\hat{\beta}_2\right) = E\left(\hat{\beta}_2\right) - \beta_2 = \frac{1}{N}$$

- But the bias converges to zero as $N \rightarrow \infty$:

Eq. 5B.3

$$\lim_{N \to \infty} \text{bias}\left(\hat{\beta}_2\right) = \lim_{N \to \infty}\left[E\left(\hat{\beta}_2\right) - \beta_2\right] = 0$$

  – The estimator is said to be **asymptotically unbiased**

- Consistency for an estimator can be established by showing that the estimator is either unbiased or asymptotically unbiased, and that its variance converges to zero as $N \to \infty$:

Eq. 5B.4

$$\lim_{N \to \infty} \operatorname{var}\left(\hat{\beta}_2\right) = 0$$

- Because the probability density function of a consistent estimator collapses around the true parameter, and the probability that an estimator b_2 will be close to the true parameter β_2 approaches one, the estimator b_2 is said to "converge in probability" to β_2
  - Notation

$$\underset{N \to \infty}{\text{plim}}\left(b_2\right) = \beta_2$$

$$b_2 \xrightarrow{p} \beta_2$$

- As $N \to \infty$ the probability density function of the standardized estimator has a distribution that approaches the standard normal:

$$\frac{b_k - \beta_k}{\sqrt{\operatorname{var}(b_k)}} \overset{a}{\sim} N(0,1)$$

  - The estimator is asymptotically normal and generally write:

$$b_k \overset{a}{\sim} N\left(\beta_k, \operatorname{var}(b_k)\right)$$

- Recall that the data generation process for the simple linear regression model is given by:

$$y_i = E(y_i \mid x_i) + e_i = \beta_1 + \beta_2 x_i + e_i, \quad i = 1, \ldots, N$$

- The Monte Carlo parameter values are $\beta_1 = 100$ and $\beta_2 = 10$
  - The value of x_i is 10 for the first $N = 2$ observations and 20 for the remaining $N = 2$ observations, so that the regression functions are:

$$E\left(y_i \mid x_i = 10\right) = 100 + 10x_i = 100 + 10 \times 10 = 200,$$

$$i = 1, \ldots, N/2$$

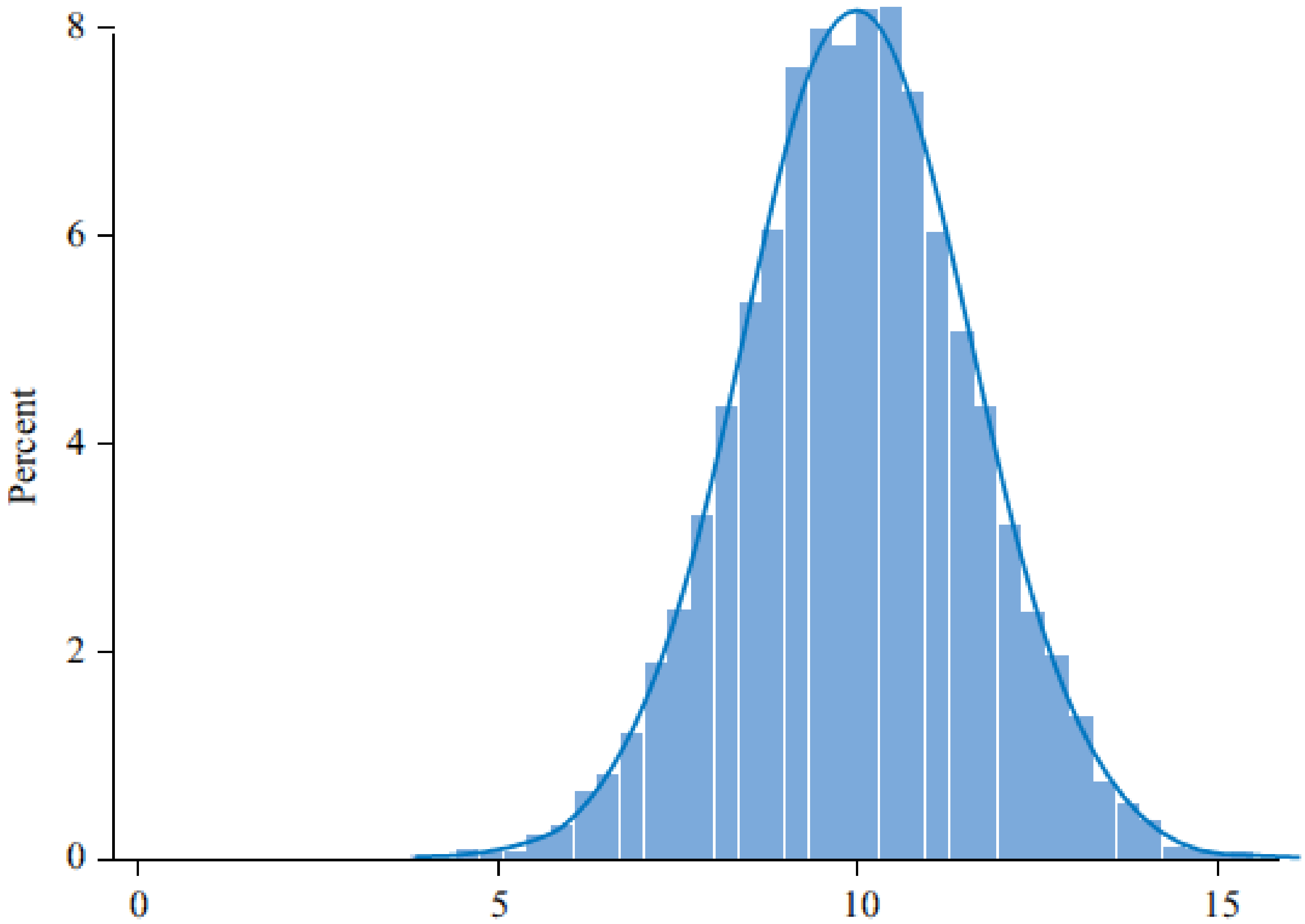
$$E\left(y_i \mid x_i = 20\right) = 100 + 20x_i = 100 + 10 \times 20 = 300,$$

$$i = \left(N/2\right) + 1, \ldots, N$$

Table 5B.1 The least squares estimators, tests, and interval estimators

N	$\overline{b}_1$	$\overline{b}_2$	$\overline{\hat{\sigma}^2}$	*REJECT*	*COVER*	*CLOSE*
20	99.4368	10.03317	2496.942	0.0512	0.9538	0.3505
40	100.0529	9.99295	2498.030	0.0524	0.9494	0.4824
100	99.7237	10.01928	2500.563	0.0518	0.9507	0.6890
200	99.8427	10.00905	2497.473	0.0521	0.9496	0.8442
500	100.0445	9.99649	2499.559	0.0464	0.9484	0.9746
1,000	100.0237	9.99730	2498.028	0.0517	0.9465	0.9980

FIGURE 5B.2 Histogram of the estimates b_2 for N = 40



- Working with nonlinear functions of the estimated parameters requires additional tools, because even if the regression errors are normal, nonlinear functions of them are not normally distributed in finite samples, and usual variance formulas do not apply

- The Taylor series approximation is:

$$f(x) \cong f(a) + \left.\frac{df(x)}{dx}\right|_{x=a}(x-a) = f(a) + f'(a)(x-a)$$

  - We apply this to $g_1(b_2)$ replacing x with b_2 and a with β_2

Eq. 5B.5

$$g_1(b_2) \cong g_1(\beta_2) + g'(\beta_2)(b_2 - \beta_2)$$

- This Taylor series expansion of shows the following:
  1. If $E(b_2) = \beta_2$ then $E[g_1(b_2)] \cong \beta_2$.
  2. If b_2 is a biased but consistent estimator, so that $b_2 \xrightarrow{p} \beta_2$, then $g_1(b_2) \xrightarrow{p} g_1(\beta_2)$
  3. The variance of $g_1(b_2)$ is given by

     $$\text{var}\left[g_1(b_2)\right] \cong \left[g_1'(\beta_2)\right]^2 \text{var}(b_2)$$

     which is known as the **delta method**.

- This Taylor series expansion of shows (Continued)
  - The delta method follows from working with the Taylor series approximation:

$$\begin{aligned}
\operatorname{var}\left[g_1(b_2)\right] &\cong \operatorname{var}\left[g_1(\beta_2) + g_1'(\beta_2)(b_2 - \beta_2)\right] \\
&= \operatorname{var}\left[g_1'(\beta_2)(b_2 - \beta_2)\right] && \text{because } g_1(\beta_2) \text{ is not random} \\
&= \left[g_1'(\beta_2)\right]^2 \operatorname{var}(b_2 - \beta_2) && \text{because } g_1'(\beta_2) \text{ is not random} \\
&= \left[g_1'(\beta_2)\right]^2 \operatorname{var}(b_2) && \text{because } \beta_2 \text{ is not random}
\end{aligned}$$

- This Taylor series expansion of shows (Continued)

  4. The estimator $g_1(\text{b}_2)$ has an approximate normal distribution in large samples:

Eq. 5B.6

$$g_1(b_2) \overset{a}{\sim} N\left[g_1(\beta_2), \left[g_1'(\beta_2)\right]^2 \operatorname{var}(b_2)\right]$$

- A fitted regression is

$$\hat{y} = 87.44311 + 10.68456x$$
$$(\text{se}) \quad (33.8764) \quad (2.1425)$$

  - Consider the nonlinear function:

$$g_1(\beta_2) = \exp(\beta_2/10)$$

  - The estimated value of the nonlinear function is:

$$g_1(b_2) = \exp(b_2/10) = \exp(10.68456/10) = 2.91088$$

Table 5B.2 Estimated covariance matrix

	b_1	b_2
b_1	1147.61330	–68.85680
b_2	–68.85680	4.59045

Table 5B.3 Estimates and tests of $g_1(\beta_2) = \exp(\beta_2) = 10$

N	$\overline{\exp(b_2/10)}$	*REJECT*
20	2.79647	0.0556
40	2.75107	0.0541
100	2.73708	0.0485
200	2.72753	0.0503
500	2.72001	0.0522
1000	2.71894	0.0555

- The estimated variance is:

$$\widehat{\text{var}}\left[g_1(b_2)\right] = \left[g_1'(b_2)\right]^2 \widehat{\text{var}}\left[b_2\right] = \left[\exp(b_2/10)\times(1/10)\right]^2 \widehat{\text{var}}\left[b_2\right]$$
$$= \left[\exp(10.68456/10)\times(1/10)\right]^2 \times 4.59045$$
$$= 0.38896$$

- se(g_1(b$_2$)) = 0.62367
- The 95% confidence interval is:

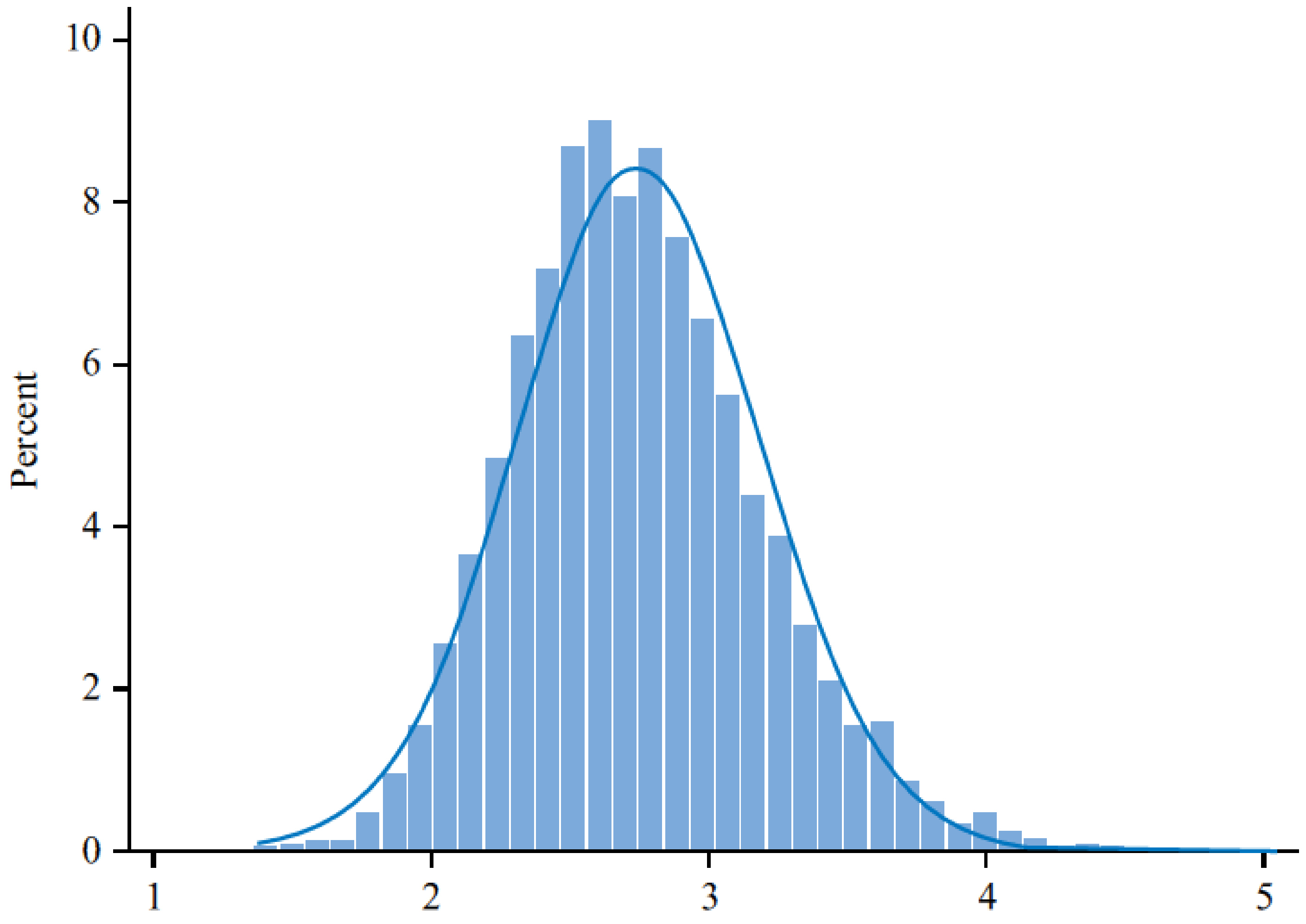
$$g_1(b_2) \pm t_{(0.975,20-2)} se\left[g_1(b_2)\right] = 2.91088 \pm 2.10092 \times 0.62367$$
$$= \left[1.60061, 4.22116\right]$$

- The test statistic is:

$$t = \frac{g_1(b_2) - 2.71828}{se\left[g_1(b_2)\right]} \sim t_{(N-2)}$$

  – The histogram of the estimates for sample size $N = 40$ in Figure 5B.3 shows only the very slightest deviation from normality, which is why the t-test performs so well.

FIGURE 5B.3 Histogram of $g_1(b_2) = \exp(b_2) = 10$



- When working with functions of two (or more) parameters, the approach is much the same, but the Taylor series approximation changes to a more general form:

Eq. 5B.7

$$g_2(b_1, b_2) \cong g_2(\beta_1, \beta_2) + \frac{\partial g_2(\beta_1, \beta_2)}{\partial \beta_1}(b_1 - \beta_1) + \frac{\partial g_2(\beta_1, \beta_2)}{\partial \beta_2}(b_2 - \beta_2)$$

- This Taylor series expansion of shows the following:
  1. If $E(b_1) = \beta_1$ and $E(b_2) = \beta_2$ then $E[g_2(b_1, b_2)] \cong g_2(\beta_1, \beta_2)$.
  2. If b_1 and b_2 are consistent estimators, so that $b_1 \xrightarrow{p} \beta_1$ and $b_2 \xrightarrow{p} \beta_2$, then $g_2(b_1, b_2) \xrightarrow{p} g_2(\beta_1, \beta_2)$

- This Taylor series expansion of shows (Continued):
  3. The variance of $g_1(\mathrm{b}_2)$ is given by:

Eq. 5B.8

$$\operatorname{var}\left[g_2(b_1,b_2)\right] \cong \left[\frac{\partial g_2(\beta_1,\beta_2)}{\partial \beta_1}\right]^2 \operatorname{var}(b_1) + \left[\frac{\partial g_2(\beta_1,\beta_2)}{\partial \beta_2}\right]^2 \operatorname{var}(b_2) + 2\left[\frac{\partial g_2(\beta_1,\beta_2)}{\partial \beta_1}\right]\left[\frac{\partial g_2(\beta_1,\beta_2)}{\partial \beta_2}\right] \operatorname{cov}(b_1,b_2)$$

- This Taylor series expansion of shows (Continued)

4. The estimator $g_2(b_1,b_2)$ has an approximate normal distribution in large samples:

Eq. 5B.9

$$g_2(b_1,b_2) \overset{a}{\sim} N\left(g_2(\beta_1,\beta_2), \operatorname{var}\left[g_2(b_1,b_2)\right]\right)$$

Table 5B.4 Estimates $g_2(b_1, b_2) = b_1/b_2$

N	$\overline{b_1/b_2}$
20	11.50533
40	10.71856
100	10.20997
200	10.10097
500	10.05755
1000	10.03070

- The nonlinear function of two parameters that we consider is $g_2(b_1, b_2) = b_1/b_2$
  - The derivatives are:

$$\frac{\partial g_2(\beta_1,\beta_2)}{\partial \beta_1} = \frac{1}{\beta_2}$$

$$\frac{\partial g_2(\beta_1,\beta_2)}{\partial \beta_2} = -\frac{\beta_1}{\beta_2^2}$$

- The estimate $g_2(b_1, b_2) = b_1/b_2$ $= 87.44311 / 10.68456 = 8.18406$ and its estimated variance is:

$$\widehat{\operatorname{var}}\left[g_2\left(b_1,b_2\right)\right]=\left[\frac{1}{b_2}\right]^2\widehat{\operatorname{var}}\left(b_1\right)+\left[-\frac{b_1}{b_2^2}\right]^2\widehat{\operatorname{var}}\left(b_2\right)+2\left[\frac{1}{b_2}\right]\left[-\frac{b_1}{b_2^2}\right]\widehat{\operatorname{cov}}\left(b_1,b_2\right)$$
$$=22.61857$$

  – The resulting 95% interval estimate for b1/b2 is [1.807712, 18.17583]

FIGURE 5B.4a Histogram of $g_2(b_1, b_2) = b_1/b_2$, $N = 40$

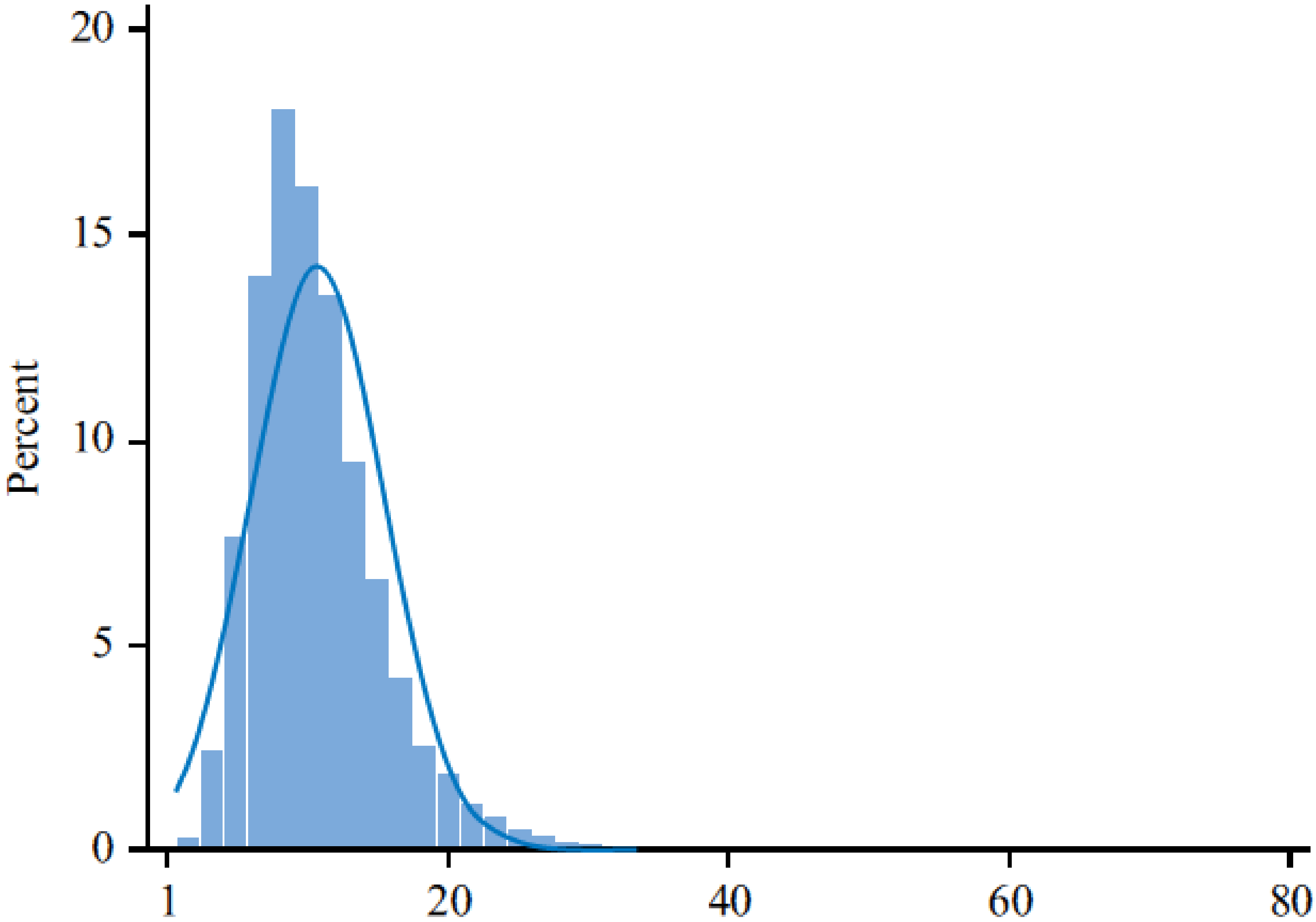


FIGURE 5B.4b Histogram of $g_2(b_1, b_2) = b_1/b_2$, $N = 200$