Estimating the Parameters of the Multiple Regression Model

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK | Gabelli School of Business

Estimating the Parameters
of the Multiple Regression Model

- We will discuss estimation in the context of the model in equation 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

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Least Squares
Estimation Procedure

- 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)

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

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

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Least Squares Estimation Procedure (cont.)

- Formulas for b_1, b_2, and b_3, obtained by minimizing equation 5.9, are estimation procedures, which are called the **least squares estimators** of the unknown parameters
  - In general, because 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 as OLS, to distinguish them from other estimators

FORDHAM | Gabelli School of Business
THE JESUIT UNIVERSITY OF NEW YORK

Estimating
the Error Variance, σ²

- We need to estimate the error variance, σ²

- Recall that $\sigma^2 = \text{var}(e_i \mid X) = E\left(e_i^2 \mid X\right) = E\left(e_i^2\right)$

- But the squared errors are unobservable, so we develop an estimator for σ² based on the squares of the least squares residuals:

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

Estimating
the Error Variance, σ² (cont.)

- An estimator for σ² that uses the information from 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

Measuring Goodness of Fit

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

FORDHAM | Gabelli School of Business

Measuring Goodness of Fit (cont.)

- 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

The Frisch–Waugh–Lovell (FWL) Theorem

- It helps us 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.

The Frisch–Waugh–Lovell (FWL) Theorem (cont.)

3. Estimate the simple regression $\widetilde{SALES}_i = \beta_3 \, \widetilde{ADVERT}_i + \widetilde{e}_i$ with no constant term. The estimate of β_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, $\widetilde{e}_i = \widetilde{SALES}_i - $ b3 $\widetilde{ADVERT}_i$. Compare these residuals to those from the complete model.

FORDHAM | Gabelli School of Business
THE JESUIT UNIVERSITY OF NEW YORK

Finite Sample Properties of the Least Squares Estimator

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

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.

FORDHAM | Gabelli School of Business
THE JESUIT UNIVERSITY OF NEW YORK

The Variances and Covariances of the Least Squares Estimators, Part I

- We can show that:

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

- Where:

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

The Variances and Covariances of the Least Squares Estimators, Part II

- 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

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK | Gabelli School of Business

The Variances and Covariances
of the Least Squares Estimators, Part III

- 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

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

FORDHAM | Gabelli School of Business
THE JESUIT UNIVERSITY OF NEW YORK

The Distribution of the Least Squares Estimators, Part I

- Consider the general form 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

FORDHAM | Gabelli School of Business
THE JESUIT UNIVERSITY OF NEW YORK

The Distribution of the Least Squares Estimators, Part II

- Because 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)

$$Z = \frac{b_k - \beta_k}{\sqrt{\mathrm{var}(b_k \mid X)}} \sim N(0,1), \qquad \text{for } k = 1, 2, \ldots, K$$

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

The Distribution of the Least Squares Estimators, Part III

- 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)

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

FORDHAM | Gabelli School of Business

The Distribution of the Least Squares Estimators, Part IV

- What happens if the errors are not normally distributed?

  - Then the least squares estimator will not be normally distributed and equations 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 equations 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

FORDHAM THE JESUIT UNIVERSITY OF NEW YORK | Gabelli School of Business

Interval Estimation

Interval Estimation for a Single Coefficient, Part I

- Suppose we are interested in finding a 95% interval estimate for β_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) $P(-t_c < t_{(72)} < t_c) = 0.95$

- Using $t_c = 1.993$, we can rewrite (5.17) as:

$$P\left(-1.993 \leq \frac{b_2 - \beta_2}{\text{se}(b_2)} \leq 1.993\right) = 0.95$$

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Interval Estimation
for a Single Coefficient, Part II

- The interval endpoints are:

  - (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 β_2

- Before the data are collected, we have confidence in the interval estimation procedure (estimator) because of its performance over all possible samples

Interval Estimation
for a Single Coefficient, Part III

- 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

Interval Estimation for a
Linear Combination of Coefficients, Part I

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:

$$\lambda = E\left(SALES_1\right) - E\left(SALES_0\right)$$

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

$$-\left[\beta_1 + \beta_2 PRICE_0 + \beta_3 ADVERT_0\right]$$

$$= -0.4\beta_2 + 0.8\beta_3$$

Interval Estimation for a
Linear Combination of Coefficients, Part II

- A point estimate would be:

$$\hat{\lambda} = -0.4b_2 + 0.8b_3 = -0.4 \times (-7.9079) + 0.8 \times 1.8626$$
$$= 4.6532$$

- A 90% interval would be:

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

FORDHAM THE JESUIT UNIVERSITY OF NEW YORK | Gabelli School of Business

Interval Estimation for a
Linear Combination of Coefficients, Part III

- Thus, a 90% interval is (4.6532 − 1.666 × 0.7096, 4.6532 + 1.666 × 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 $3,471 and $5,835

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Hypothesis Testing

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 α and determine the rejection region
4. Calculate the sample value of the test statistic and, if desired, the p-value
5. State your conclusion

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK | Gabelli School of Business

Testing the Significance of a Single Coefficient

- 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

Testing the Significance
of a Single Coefficient (cont.)

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

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)

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

One-Tail Hypothesis Testing for a Single Coefficient, Part II

- 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:

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

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

One-Tail Hypothesis
Testing for a Single Coefficient, Part III

- The test statistic is:

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

- And the p-value is:

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

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

Hypothesis Testing for a Linear Combination of Coefficients

Example 5.12: testing the effect of changes in price and advertising

- The null and alternative hypotheses are:

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

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

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

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

Hypothesis Testing for a Linear Combination of Coefficients (cont.)

- At a 5% significance level, we reject H_0 if t ≥ 1.666 or if the *p*-value ≤ 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:

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

- Because 1.622 < 1.666, we do not reject H_0

Nonlinear Relationships

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: Cost and Product Curves, Part I

- Consider the average cost equation:

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

- And the total cost function:

$$TC = \alpha_1 + \alpha_2 Q + \alpha_3 Q^2 + \alpha_4 Q^3 + e$$
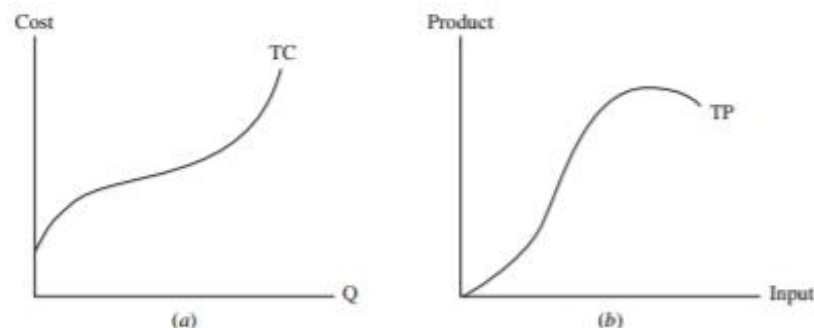
**FIGURE 5.2** (*a*) Total cost curve and (*b*) total product curve.
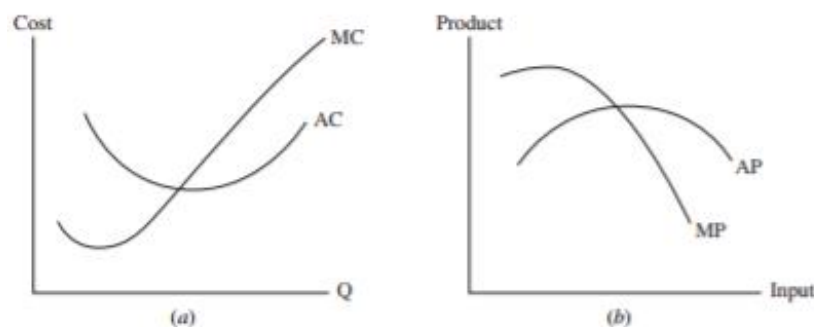
FIGURE 5.3 Average and marginal (*a*) cost curves and (*b*) product curves.

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Example: Cost and Product Curves, Part II

- 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:

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

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Example: Cost and Product Curves, Part III

- 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 $\alpha_4 > 0$

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Large Sample Properties of the Least Squares Estimator

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 = 1,000 or larger could be necessary.

Consistency

- 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 → ∞. Or, using the concept of a limit, the estimator b2 is consistent if
$$\lim_{n \to \infty} P\,(\beta_2 - \varepsilon \leq b_1 \leq \beta_2 + \varepsilon) = 1.$$

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Consistency (cont.)

- As the sample size increases, the probability density function (*pdf*) becomes narrower.

- As the sample size changes, the center of the pdfs remains at β2. However, as the sample size N gets larger, the variance of the estimator b2 becomes smaller.

- The probability that b2 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 b2 falling within the limits increases toward 1.

FORDHAM | Gabelli School of Business
THE JESUIT UNIVERSITY OF NEW YORK
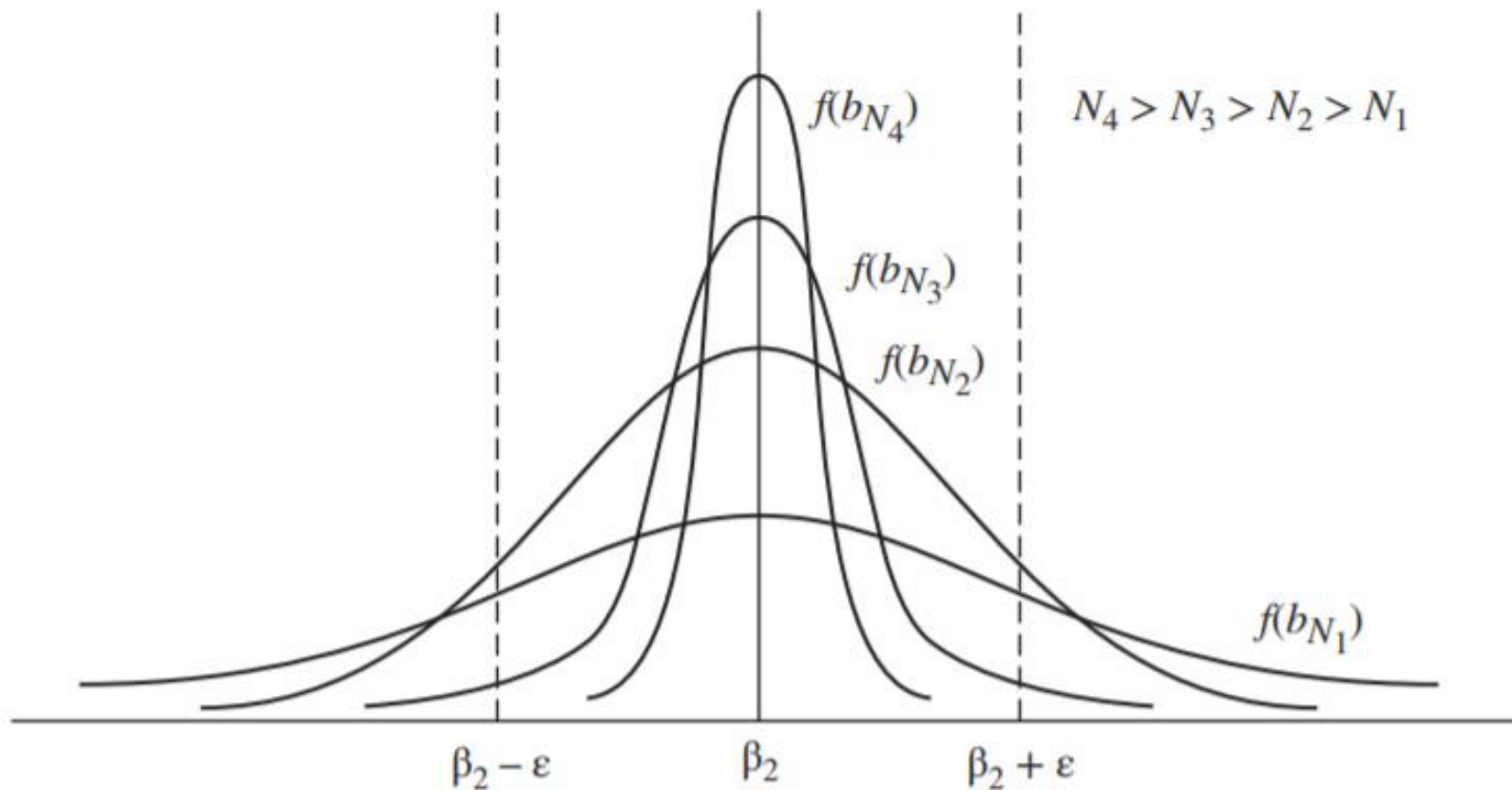
FIGURE 5.5 An illustration of consistency.

Asymptotic Normality, Part I

- 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 $\sqrt{N}\text{b}_2$

- It follows that $E(\sqrt{N}\text{b}_2) = \sqrt{N}\beta_2$

- And (5.35) $\sqrt{N}\text{b}_2 \sim (\sqrt{N}\beta_2, \text{C}_x)$

Asymptotic Normality, Part II

- Applying a central limit theorem to the $\sum_{i=1}^{N}(x_i - \bar{x})e_i/\sqrt{N}$ and using $[s_x^2]^{-1} \xrightarrow{P} Cx$, it can be shown that the statistic obtained by normalizing equation 5.35 so that it has mean 0 and variance 1, will be approximately normally distributed:

$$\frac{\sqrt{N}(b2 - \beta2)}{\sqrt{\sigma^2 Cx}} \, N(0,1)$$

- 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

Asymptotic Normality, Part III

- Thus, we can write equation 5.36 as: $t = \dfrac{\sqrt{N}(b2 - \beta2)}{\sqrt{\widehat{\sigma^2}/s_x^2}}$ N(0,1)

- This is precisely the t-statistic that we use for interval estimation and hypothesis testing; the result in equation 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 in the simple regression model; the result also holds for estimators of the coefficients in the multiple regression model

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Relaxing Assumptions, Part I

- **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 equation 5.38 means we can no longer show that the least squares estimator is unbiased

Relaxing Assumptions, Part II

- **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, so it is no longer reasonable to assume they are independent

  - The likely violation of $cov(e_t, x_s t) = 0$ 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

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Relaxing Assumptions, Part III

- 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 $cov(e_t, x_{tk}) = 0$ for t = 1, 2,…, N; k = 1, 2,…, K

- Errors and the explanatory variables that satisfy equation 5.44 are said to be contemporaneously uncorrelated

- We do not insist that $cov(e_t, x_{tk}) = 0$ for t ≠ s

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 is in the next example.

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business

Example: The Optimal Level of Advertising

- 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: The Optimal Level of Advertising (cont.)

- 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 \$1,757 and \$2,271

FORDHAM | Gabelli School of Business
THE JESUIT UNIVERSITY OF NEW YORK

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
- 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

FORDHAM | Gabelli School
THE JESUIT UNIVERSITY OF NEW YORK | of Business