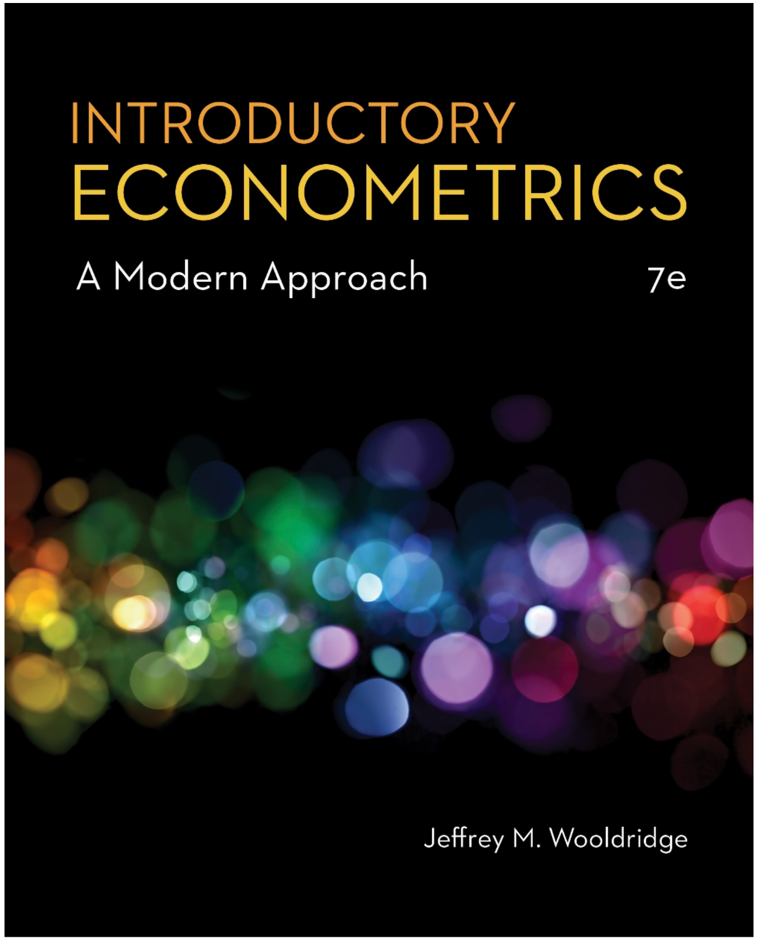
Chapter 5

Multiple Regression Analysis: OLS Asymptotics



Multiple Regression Analysis: OLS Asymptotics (1 of 7)

- So far we focused on properties of OLS that hold for any sample

- Properties of OLS that hold for any sample/sample size
  - Expected values/unbiasedness under MLR.1 – MLR.4
  - Variance formulas under MLR.1 – MLR.5
  - Gauss-Markov Theorem under MLR.1 – MLR.5
  - Exact sampling distributions/tests under MLR.1 – MLR.6

- Properties of OLS that hold in large samples
  - Consistency under MLR.1 – MLR.4
  - Asymptotic normality/tests under MLR.1 – MLR.5
    - Note that we drop MLR.6

The image part with relationship ID rId18 was not found in the file.

Multiple Regression Analysis: OLS Asymptotics (2 of 7)

- **Consistency**

An estimator θ_n is consistent for a population parameter θ if

$$P\left(|\theta_n - \theta| < \epsilon\right) \rightarrow 1 \text{ for arbitrary } \epsilon > 0 \text{ and } n \rightarrow \infty.$$

Alternative notation: $plim\ \theta_n = \theta$ ← The estimate converges in probability to the true population value

- Interpretation:
  - Consistency means that the probability that the estimate is arbitrarily close to the true population value can be made arbitrarily high by increasing the sample size
- Consistency is a minimum requirement for sensible estimators

Multiple Regression Analysis: OLS Asymptotics (3 of 7)

- **Theorem 5.1 (Consistency of OLS)**

$$MLR.1 - MLR.4 \quad \Rightarrow \quad plim\ \widehat{\beta}_j = \beta_j, \quad j = 0, 1, \ldots, k$$

- Special case of simple regression model

$$plim\ \widehat{\beta}_1 = \beta_1 + Cov(x_1, u)/Var(x_1)$$

One can see that the slope estimate is consistent if the explanatory variable is exogenous, i.e. un-correlated with the error term.

- **Assumption MLR.4'**

$$E(u) = 0$$

$$Cov(x_j, u) = 0$$

All explanatory variables must be uncorrelated with the error term. This assumption is weaker than the zero conditional mean assumption MLR.4.

The image part with relationship ID rId18 was not found in the file.

Multiple Regression Analysis: OLS Asymptotics (4 of 7)

- For consistency of OLS, only the weaker MLR.4 is needed
- Asymptotic analog of omitted variable bias

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v$$ ← True model

$$y = \beta_0 + \beta_1 x_1 + [\beta_2 x_2 + v] = \beta_0 + \beta_1 x_1 + u$$ ← Misspecified model

$$\Rightarrow \quad plim\ \tilde{\beta}_1 = \beta_1 + Cov(x_1, u)/Var(x_1)$$

$$= \beta_1 + \beta_2 Cov(x_1, x_2)/Var(x_1) = \beta_1 + \beta_2 \delta_1$$ ← Bias

There is no omitted variable bias if the omitted variable is irrelevant or uncorrelated with the included variable

Multiple Regression Analysis: OLS Asymptotics (5 of 7)

- **Asymptotic normality and large sample inference**
  - In practice, the normality assumption MLR.6 is often questionable
  - If MLR.6 does not hold, the results of t- or F-tests may be wrong
  - Fortunately, F- and t-tests still work if the sample size is large enough
  - Also, OLS estimates are normal in large samples even without MLR.6

- **Theorem 5.2 (Asymptotic normality of OLS)**
  - Under assumptions MLR.1 – MLR.5:

$$\frac{(\widehat{\beta}_j - \beta_j)}{se(\widehat{\beta}_j)} \overset{a}{\sim} \text{Normal}(0, 1)$$

← In large samples, the standardized estimates are normally distributed

also $plim\ \widehat{\sigma}^2 = \sigma^2$

The image part with relationship ID rId18 was not found in the file.

Multiple Regression Analysis: OLS Asymptotics (6 of 7)

- **Practical consequences**
  - In large samples, the t-distribution is close to the Normal (0,1) distribution
  - As a consequence, t-tests are valid in large samples without MLR.6
  - The same is true for confidence intervals and F-tests
  - Important: MLR.1 – MLR.5 are still necessary, esp. homoskedasticity

- **Asymptotic analysis of the OLS sampling errors**

$$\widehat{Var}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}$$

Converges to σ^2

Converges to $n \cdot Var(x_j)$

Converges to a fixed number

The image part with relationship ID rId18 was not found in the file.

Multiple Regression Analysis: OLS Asymptotics (7 of 7)

- **Asymptotic analysis of the OLS sampling errors (cont.)**

$\widehat{Var}(\widehat{\beta}_j)$ shrinks with the rate $1/n$

$se(\widehat{\beta}_j)$ shrinks with the rate $\sqrt{1/n}$

- This is why large samples are better
- Example: Standard errors in a birth weight equation

$$n = 1,388 \;\Rightarrow se(\widehat{\beta}_{cigs}) = .00086$$

$$n = 694 \;\Rightarrow se(\widehat{\beta}_{cigs}) = .0013$$

$$\frac{.00086}{.0013} \approx \sqrt{\frac{694}{1,388}}$$

Use only the first half of observations