AFRE 835: Introductory Econometrics

Chapter 5: Multiple Regression Analysis: OLS Asymptotics

Spring 2017

Introduction

- In chapter 3 we derived several *finite sample* properties of the OLS estimator, characterizing its conditional mean and variance under the Gauss-Markov assumptions (MLR.1 through MLR.5).
- Chapter 4 added MLR.6 (normality of the errors) in order to fully characterized the finite sampling distribution of the OLS estimator.
  - ... which in turn let us construct t and F statistics for use in hypothesis testing.
- The problem is that MLR.6 is a very strong assumption and unlikely to hold in many settings.
- Fortunately, there is an alternative to MLR.6, relying on the *asymptotic (or large sample)* properties of estimators and test statistics.
- These properties suggest that, with sufficiently large sample sizes, the t and F statistics described in chapter 4 are *approximately* correct.
- Here, we will only quickly review the basic results of this chapter.

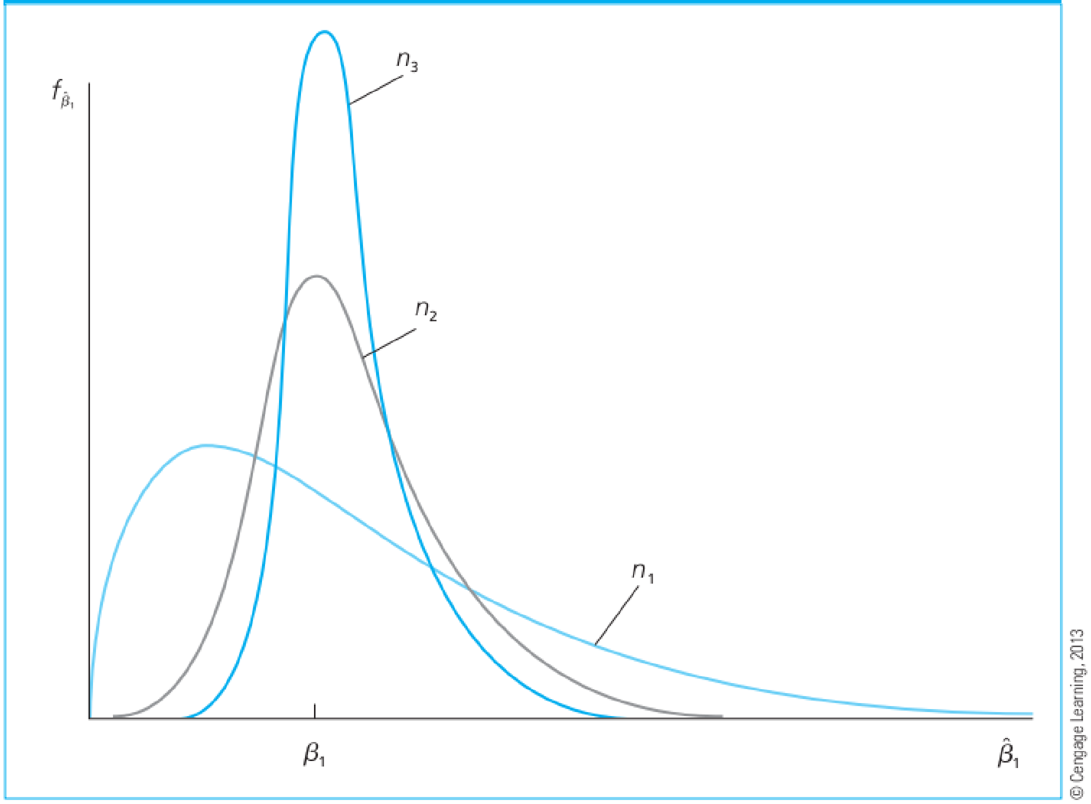
Outline

1. Consistency
2. Asymptotic Normality

Consistency

- **Consistency:** Lrt W_n be an estimator of θ based on a sample $Y_1, \ldots, Y_n$ of size n. Then W_n is a **consistent estimator** of θ if for every $\epsilon > 0$,
$$Pr(|W_n - \theta| > \epsilon) \to 0 \text{ as } n \to \infty. \quad (1)$$
- Intuitively, consistency simply says that, as the sample size increases, the sampling distribution of W_n collapses around θ.
- Note: Consistency is a weaker condition than unbiasedness.
- **Theorem 5.1 (Consistency of OLS):** Under Assumptions MLR.1 through MLR.4, the OLS estimator $\hat{\beta}_j$ is consistent for β_j $j = 0, \ldots, k$.

FIGURE 5.1 Sampling distributions of $\hat{\beta}_1$ for sample sizes $n_1 < n_2 < n_3$.



Asymptotic Normality

- While consistency is an important attribute of an estimator, it is not enough to support inference.
- In chapter 4 we relied on the normality of the error term u to justify finite sample inference.
- **Asymptotic Normality**: Let $\{Z_n : n = 1, 2, \dots\}$ be a sequence of random variables, such that for all numbers z:

$$P(Z_n \leq z) \rightarrow \Phi(z) \text{ as } n \rightarrow \infty, \tag{2}$$

where $\Phi(z)$ is the standard normal cdf. Then Z_n is said to have an *asymptotic normal distribution*, often denoted as $Z_n \overset{a}{\sim} \mathcal{N}(0,1)$.

Asymptotic Normality (cont'd)

- **Theorem 5.2 (Asymptotic Normality of OLS)**: Under the Gauss-Markov Assumptions MLR.1 through MLR.5,
  i. $\sqrt{n}(\hat{\beta}_j - \beta_j) \stackrel{a}{\sim} \mathcal{N}(0, \frac{\sigma^2}{a_j^2})$, where $\frac{\sigma^2}{a_j^2}$ is the **asymptotic variance** of $\sqrt{n}(\hat{\beta}_j - \beta_j)$ for the slope coefficients, $a_j = plim(\frac{1}{n}\sum_{i=1}^{n} \hat{r}_{ij}^2)$, where $\hat{r}_{ij}^2$ are the residuals from regressing x_j on the other dependent variables). We say that $\hat{\beta}_j$ is asymptotically normally distributed.
  ii. $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2 = Var(u)$.
  iii. For each j
  $$\frac{(\hat{\beta}_j - \beta_j)}{sd(\hat{\beta}_j)} \stackrel{a}{\sim} \mathcal{N}(0,1) \tag{3}$$
  and
  $$\frac{(\hat{\beta}_j - \beta_j)}{se(\hat{\beta}_j)} \stackrel{a}{\sim} \mathcal{N}(0,1) \tag{4}$$
  where $se(\hat{\beta}_j)$ is the usual OLS standard error.