ECON 5340 Class Notes
Chapter 5. Large-Sample Properties of the LS Estimator

1 Introduction

Unlike the simple linear regression model, in many cases we cannot calculate the exact distribution of our estimators. This is generally true when we relax Classical assumption #6, which we do here. Fortunately, however, we can often calculate approximate distributions that hold when the sample size is large. This is the focus of chapter 5.

2 Consistency of b

Recall, a **consistent estimator** has the following property

$$\lim_{n\to\infty} \Pr(|b-\beta| < \delta) = 1$$

for any positive δ. It is said that the probability limit of b is β, that is $plim(b) = \beta$. Next, we are going to establish the consistency of b.

Continue to assume that X is nonstochastic and

$$\lim_{n\to\infty} \frac{1}{n}(X'X) = Q,$$

is a positive-definite finite matrix. This condition is fairly restrictive (less restrictive assumptions can be used) and guarantees that the explanatory data are "well-behaved" in the sense that their variance does not get too large. Here is a counter-example.

- Counterexample. Consider the time-series model

$$y_t = \beta_1 + \beta_2 t + \epsilon_t$$

where $t = 1, \ldots, n$. In this case,

$$X'X = \begin{bmatrix} n & \sum_{t=1}^{n} t \\ \sum_{t=1}^{n} t & \sum_{t=1}^{n} t^2 \end{bmatrix} = \begin{bmatrix} n & \frac{n(n+1)}{2} \\ \frac{n(n+1)}{2} & \frac{n(n+1)(2n+1)}{6} \end{bmatrix} \Longrightarrow \lim_{n\to\infty} \frac{1}{n}(X'X) = \begin{bmatrix} 1 & \infty \\ \infty & \infty \end{bmatrix}.$$

To show consistency, rewrite b as

$$b = \beta + (\frac{1}{n}X'X)^{-1}(\frac{1}{n}X'\epsilon).$$

Taking the probability limit gives

$$\begin{aligned} plim(b-\beta) &= plim(\frac{1}{n}X'X)^{-1} plim(\frac{1}{n}X'\epsilon) \\ &= (plim(\frac{1}{n}X'X))^{-1} plim(\frac{1}{n}X'\epsilon) \\ &= Q^{-1} \times 0 = 0 \end{aligned}$$

where $plim(\frac{1}{n}X'X)^{-1} = (plim(\frac{1}{n}X'X))^{-1}$ via Slutsky's Theorem (Greene Theorem D.12) and $plim(\frac{1}{n}X'\epsilon) = 0$ because $\frac{1}{n}X'\epsilon$ converges in mean square to zero (Greene Theorem D.11). As a result, $plim(b) = \beta$ or b is a consistent estimator of β.

3 Asymptotic Distribution of b

Continue to assume that X is nonstochastic, $\lim_{n\to\infty} \frac{1}{n}(X'X) = Q$ and $\epsilon \sim (0, \sigma^2 I)$. Because b is a consistent estimator of β, the limiting distribution of b is degenerate (i.e., a spike at β). However, using the Central Limit Theorem, we can take a stabilizing transformation of b to produce a non-degenerate limiting distribution

$$\sqrt{n}(b-\beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}).$$

This result suggests that, in large samples, we can approximate the distribution of b as $N(\beta, \frac{\sigma^2}{n}Q^{-1})$. We call this the **asymptotic distribution of b** or $b \overset{asy}{\sim} N(\beta, \frac{\sigma^2}{n}Q^{-1})$. A few notes.

- If $f(b)$ is continuous, $f(b) \overset{asy}{\sim} N(f(\beta), \Gamma\frac{\sigma^2}{n}Q^{-1}\Gamma')$ where $\Gamma = \frac{\partial f(\beta)}{\partial \beta}$.
- If we add that $\epsilon \sim N(0, \sigma^2 I)$, then b has the exact distribution $N(\beta, \sigma^2(X'X)^{-1})$.
- Since σ^2 and Q are unknown, we can replace them with s^2 and $\frac{1}{n}(X'X)$ respectively. This produces our

standard estimate of $var(b)$, $s^2(X'X)^{-1}$, which is also a consistent estimate of the unknown asymptotic variance.

4 Asymptotic Behavior of Test Statistics

Earlier we assumed that the errors were normal (i.e., $\epsilon \sim N(0, \sigma^2 I)$), and showed that

- $t = (b - \beta)/s_b \sim t(n-k)$.
- $F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k)$.

If instead we did not impose normality (i.e., $\epsilon \sim (0, \sigma^2 I)$), then we can show

- $t = (b - \beta)/s_b \overset{asy}{\sim} N(0, 1)$.
- $(k-1)F = \frac{R^2}{(1-R^2)/(n-k)} \overset{asy}{\sim} \chi^2(k-1)$.

5 Instrumental Variables and 2SLS

Covered later.

6 Measurement Error

Covered later.

7 Normally Distributed Disturbances

Now we will reinstate the assumption that $\epsilon \sim N(0, \sigma^2 I)$.

7.1 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is an alternative estimation criterion to least squares. The principle of MLE is to select the parameters of the model so as to maximize the likelihood (or probability) that the data were generated by the model.

Given the model $Y = X\beta + \epsilon$ and the assumption on the errors above, we can write the joint probability or likelihood function as

$$\begin{aligned} L &= (2\pi\sigma^2)^{-n/2} \exp(-\epsilon'\epsilon/2\sigma^2) \\ &= (2\pi\sigma^2)^{-n/2} \exp(-(Y - X\beta)'(Y - X\beta)/2\sigma^2). \end{aligned}$$

We could maximize this likelihood directly by choosing the unknown parameters (β and σ^2), however, it is often easier to work with the log likelihood function

$$\ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta).$$

Taking derivatives with respect to β and σ^2 and setting equal to zero gives

$$\begin{aligned} \frac{\partial \ln(L)}{\partial \beta} &= \frac{1}{\sigma^2}X'(Y - X\beta) = 0 \\ \frac{\partial \ln(L)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)'(Y - X\beta) = 0. \end{aligned}$$

Solving these equations jointly for β and σ^2 produces the ML estimates

$$\begin{aligned} \hat{\beta}_{ML} &= (X'X)^{-1}X'Y = b \\ \hat{\sigma}^2_{ML} &= e'e/n = s^2\frac{n-k}{n}. \end{aligned}$$

From our earlier LS analysis, we know that $\hat{\beta}_{ML}$ is the best linear unbiased estimator of β and $\hat{\sigma}^2_{ML}$ is a biased estimate of σ^2 (recall s^2 is unbiased). All ML estimates (subject to some weak regularity conditions) have the following properties:

- Consistency.
- Asymptotic efficiency (i.e., among all consistent estimates they have the smallest asymptotic variance).
- Asymptotic normality.
- Invariance (i.e., if $\hat{\theta}_{ML}$ is the ML estimate of θ, then $g(\hat{\theta}_{ML})$ is the ML estimate of $g(\theta)$ for continuous g).

One potential drawback of MLE is that it requires the user to know the distribution of the errors, which is often assumed normal. Fortunately, there are tests based on the skewness and kurtosis of the residuals (e.g., Bera-Jarque test), that allow one to test this assumption.

7.2 Wald, Lagrange Multiplier, and Likelihood Ratio Tests

The standard t and F tests require the errors to be normally distributed. If the errors are not normally distributed, however, we can rely on the CLT and the fact that b is asymptotically normal. Assume we wish to test the (possibly nonlinear) null hypothesis, H_0: $g(\beta) = 0$. Below are three tests that will generally give different answers in small samples but are asymptotically equivalent.

- Wald Test.

$$W = g(b)' \left\{ G(b)[s^2(X'X)^{-1}]G(b)' \right\}^{-1} g(b) \overset{asy}{\sim} \chi^2(J)$$

  where $G(b) = \partial g(b)/\partial b'$ and J is the number of restrictions in the null.

- Likelihood Ratio Test.

$$\begin{aligned} LR &= -2(\ln L_* - \ln L) \\ &= n(\ln(e_*'e_*) - \ln(e'e)) \overset{asy}{\sim} \chi^2(J) \end{aligned}$$

  where an asterisk indicates the restricted (null hypothesis imposed) likelihood and residuals.

- Lagrange Multiplier Test.

$$\begin{aligned} LM &= W/(1 + W/n) \\ &= nR_0^2 \overset{asy}{\sim} \chi^2(J) \end{aligned}$$

  where R_0^2 is the coefficient of determination from a regression of e_* on X.