Understanding OLS: Asymptotic Properties

Le Wang

Motivation

As we have seen, without making further assumptions on conditional distribution of the unobservables, ϵ, we cannot dervie the distribution of $\widehat{\beta}$.

Knowing the first two moments is certainly useful. It is not enough when it comes to things such as hypothesis testing (another topic later).

It turns out that while we cannot find the exact distribution, we are able to approximate the distribution when the sample size is large.

Motivation

The theory of such approximation is called **asymptotic theory**. In what follows, we will derive the asymptotic (approximation) distribution of our OLS estimator.

Main Result: Under previous assumptions, OLS estimator is **asymptotically**

1. **Consistent** (hitting the right target; large-sample unbiasedness)
2. **Normally distributed**.

Notations and Math Preliminaries

Intuition:

1. **Consistency**: as we get more and more data, we eventually know the truth
2. **Asymptotic normality**: as we get more and more data, averages of random variables behave like normally distributed random variables

Preview of Where we are going

$$
\begin{aligned}
\widehat{\beta} &= (X'X)^{-1}X'Y \\
&= \beta + (X'X)^{-1}X'\epsilon \\
&= \beta + (\frac{1}{N}\sum x_i x_i')^{-1}(\frac{1}{N}\sum x_i\epsilon_i)
\end{aligned}
$$

Note that both denominator and numerator are sample averages of some variables

$$
\begin{aligned}
\frac{1}{N}\sum x_i x_i' &: \text{sample averages of } x_i x_i' \\
\frac{1}{N}\sum x_i \epsilon_i &: \text{sample averages of } x_i \epsilon_i
\end{aligned}
$$

Probability Theory Tools

The probability theory tools (theorems) for establishing **consistency** of estimators are

- Laws of Large Numbers (LLNs).

The tools (theorems) for establishing **asymptotic normality** are

- Central Limit Theorems (CLTs).

A comprehensive reference is White (1994), Asymptotic Theory for Econometricians, Academic Press.

Notations and Math Preliminaries

Consistent Esimator (Asymptotic Version of "Hitting the Right Target")

$\widehat{\theta}$ is consistent for θ if every element, $\widehat{\theta}_k$ **converge in probability** to θ_k. We write it as follows

$$\widehat{\theta} \xrightarrow{p} \theta$$

Notation (Skip)

Greene, Appendix D.2

Convergence in Probability

The random variable x_N **converges in probabiity** to a constant c if for any $\epsilon > 0$

$$\lim_{N \to \infty} \Pr[|x_N - c| > \epsilon] = 0$$

Law of Large Number (LLN)

Multivariate Version of Khinchine's Weak LLN (Greene, p1070 Appendix D.2)

Let $X_1, X_2, \ldots, X_N$ denote an independent and identically distributed random sample with $\mathbb{E}[X_i] = \mu$, then

$$\bar{X} = \frac{1}{N} \sum X_i \xrightarrow{p} \mathbb{E}[X_i] = \mu$$

All consistency proofs are based on a particular LLN. A LLN is a result that states the conditions under which a sample average of random variables converges to a population expectation. There are many LLN results. The most straightforward is the LLN due to Chebychev.

Slutsky's Theorem #1 (Complete Version)

Let $\{Y_N\}$ and $\{Z_N\}$ be a sequences of random variables and let b, c and d be constants.

1. If $Y_N \xrightarrow{p} c$, then $bY_N \xrightarrow{p} bc$.
2. If $Y_N \xrightarrow{p} c$ and $Z_N \xrightarrow{p} d$, then $Y_N + Z_N \xrightarrow{p} c + d$
3. If $Y_N \xrightarrow{p} c$ and $Z_N \xrightarrow{p} d$, then $\frac{Y_N}{Z_N} \xrightarrow{p} \frac{c}{d}$. provided $d \neq 0$.
   $Y_N Z_N \xrightarrow{p} cd$.
4. If $Y_N \xrightarrow{p} c$ and $h(\cdot)$ is a continuous function at c then $h(Y_N) \xrightarrow{p} h(c)$.

LLN and Consistency of OLS

$$\widehat{\beta} \xrightarrow{p} \beta$$

$$\begin{aligned}
\widehat{\beta} &= (X'X)^{-1}X'Y \\
&= \beta + (X'X)^{-1}X'\epsilon \\
&= \underbrace{\beta}_{\xrightarrow{p}\beta} + (\underbrace{\frac{1}{N}\sum x_i x_i'}_{\xrightarrow{p}\mathbb{E}[xx']})^{-1}(\underbrace{\frac{1}{N}\sum x_i \epsilon_i}_{\xrightarrow{p}\mathbb{E}[x\epsilon]=0})
\end{aligned}$$

$$\widehat{\beta} \xrightarrow{p} \beta + \mathbb{E}[xx']^{-1}\mathbb{E}[x\epsilon] = \beta$$

Convergence in Distribution

If $F_{Y_N}(y) = \Pr[Y_N \leq y] \to F_W(y) = \Pr[W \leq y]$ as $n \to \infty$ for every continuity point of the CDF of W,

$$Y_N \xrightarrow{d} W$$

Convergence in Distribution

If $F_{Y_N}(y) = \Pr[Y_N \leq y] \rightarrow F_W(y) = \Pr[W \leq y]$ as $n \rightarrow \infty$ for every continuity point of the CDF of W,

$$Y_N \xrightarrow{d} W$$

We often write out the distribution for W, for example,

$$Y_N \overset{d}{\sim} N(\mu, \sigma^2)$$

Notation (skip)

Greene, Appendix D.2.5

Convergence in Distribution

x_N converges in distribution to a random variable x with CDF $F(x)$ if for all continuity points of $F(x)$

$$\lim_{N\to\infty} |F(x_N) - F(x)| = 0$$

Several Rules

Cramer-Wold Device (Greene, p.1078)

If $Y_N \stackrel{d}{\to} X$, then $c'Y_N \stackrel{d}{\to} c'W$ for all conformable vectors c with real valued elements.

Other Rules (Greene, p.1077)

1. If $X_N \xrightarrow{d} X$ and $Y_N \xrightarrow{p} c$, then the following holds

$$X_N Y_N \xrightarrow{d} cX$$

$$X_N + Y_N \xrightarrow{d} c + X$$

$$\frac{X_N}{Y_N} \xrightarrow{d} \frac{X}{c}$$

2. If $X_N \xrightarrow{d} X$ and $h(\cdot)$ is a continuous function, then

$$h(X_N) \xrightarrow{d} h(X)$$

Convergence in Distribution (Refresher)

Remember that we have looked at some of the resuilts!

For example,

$$F_{1,N} \xrightarrow{d} \chi_1^2$$

$$t_N^2 \xrightarrow{d} N(0,1)$$

Central Limit Theorem (CLT)

Multivariate Lindeberg-Levy Central Limit Theorem (Greene, p.1082)

Let $X_1, X_2, \ldots, X_N$ denote an independent and identically distributed random sample with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \mathbb{E}[(X_i - \mu)(X_i - \mu)'] = \Sigma$, where Σ is nonsingular, then

$$\sqrt{N}(\bar{X} - \mu) \overset{d}{\sim} N(0, \Sigma)$$

Central Limit Theorem (CLT)

$$\sqrt{N}(\bar{X} - \mu) \overset{d}{\sim} N(0, \Sigma)$$

Equivalently,

$$\bar{X} \overset{d}{\sim} N(\mu, \frac{1}{N}\Sigma)$$

where the asymptotic variance of $\bar{X} = \frac{1}{N}\Sigma$.

CLT and Asymptotic Normality of OLS (main results)

Under our assumptions, we can have the following result

$$\sqrt{N}(\widehat{\beta} - \beta) \sim N(0, D\Sigma D')$$

But we need the following in practice:

$$\widehat{\beta} \sim N(\beta, \frac{1}{N} D\Sigma D')$$

CLT and Asymptotic Normality of OLS

$$\widehat{\beta} = \beta + (\frac{1}{N}\sum x_i x_i')^{-1}(\frac{1}{N}\sum x_i \epsilon_i)$$

$$\widehat{\beta} - \beta = (\frac{1}{N}\sum x_i x_i')^{-1}(\frac{1}{N}\sum x_i \epsilon_i)$$

$$\sqrt{N}(\widehat{\beta} - \beta) = \sqrt{N}\left[(\frac{1}{N}\sum x_i x_i')^{-1}(\frac{1}{N}\sum x_i \epsilon_i)\right]$$

$$\sqrt{N}(\widehat{\beta} - \beta) = \left[(\frac{1}{N}\sum x_i x_i')^{-1}(\sqrt{N}\cdot\frac{1}{N}\sum x_i \epsilon_i)\right]$$

CLT and Asymptotic Normality of OLS

$$\sqrt{N}(\widehat{\beta} - \beta) = \left[(\frac{1}{N} \sum x_i x_i')^{-1} (\sqrt{N} \cdot \frac{1}{N} \sum x_i \epsilon_i) \right]$$

$$\sqrt{N}(\widehat{\beta} - \beta) = \left(\frac{1}{N} \sum x_i x_i' \right)^{-1} \cdot \left[\sqrt{N} \cdot (\frac{1}{N} \sum x_i \epsilon_i - 0) \right]$$

$$\sqrt{N}(\widehat{\beta} - \beta) = \underbrace{\left(\frac{1}{N} \sum x_i x_i' \right)^{-1}}_{\xrightarrow{p} D \equiv \mathbb{E}[xx']^{-1}} \cdot \underbrace{\left[\sqrt{N} \cdot (\frac{1}{N} \sum x_i \epsilon_i - \mathbb{E}[x\epsilon]) \right]}_{\xrightarrow{d} N(0,\Sigma)}$$

From what we have learned about the properties of multivariate normal:

$$\sqrt{N}(\widehat{\beta} - \beta) \sim N(0, D\Sigma D')$$

Asymptotic Normality of OLS

$$\sqrt{N}(\widehat{\beta} - \beta) \sim N(0, D\Sigma D')$$

$$\Sigma = \mathbb{E}[(x\epsilon)(x\epsilon)'] = \mathbb{E}[xx'\epsilon^2]$$
$$D = \mathbb{E}[xx']^{-1}$$

Asymptotic Normality of OLS

Theoretical Result

$$\sqrt{N}(\widehat{\beta} - \beta) \sim N(0, D\Sigma D')$$

What we need:

$$\widehat{\beta} \sim N(\beta, \frac{1}{N} D\Sigma D')$$

Standard Errors of β_k is given by $\sqrt{V_{kk}}$, where

$$V = \frac{1}{N} D\Sigma D'$$

Asymptotic Normality of OLS (A Special Case)

Note that

$$\mathbb{E}[xx'\epsilon^2] = \mathbb{E}[xx'\mathbb{E}[\epsilon^2|x]] \overset{A4}{=} \mathbb{E}[xx']\sigma^2$$

$$\begin{aligned} D\Sigma D' &= \left(\mathbb{E}[xx']^{-1}\right)\left(\mathbb{E}[xx'\epsilon^2]\right)\left(\mathbb{E}[xx']^{-1}\right)' \\ &= \left(\mathbb{E}[xx']^{-1}\right)\left(\mathbb{E}[xx']\sigma^2\right)\left(\mathbb{E}[xx']^{-1}\right)' \\ &= \sigma^2\mathbb{E}[xx']^{-1} \end{aligned}$$