MA Advanced Econometrics: Properties of Least Squares Estimators

Karl Whelan

School of Economics, UCD

February 15, 2011

Part I

Least Squares: Some Finite-Sample Results

Univariate Regression Model with Fixed Regressors

- Consider the simple regression model

$$y_i = \beta x_i + \epsilon_i \tag{1}$$

  where ϵ_i ($i = 1, 2, ...n$) are i.i.d. and of mean zero and with finite variance.

- The usual textbook assumption is that x_i is a fixed deterministic regressor independent of all values of ϵ_i.

- What does a fixed regressor actually mean? It means that we are to think of x_i not as an outcome of a random process but merely as a fixed set of numbers.

- The OLS estimator is

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i \left(\beta x_i + \epsilon_i\right)}{\sum_{i=1}^{n} x_i^2} = \beta + \frac{\sum_{i=1}^{n} x_i \epsilon_i}{\sum_{i=1}^{n} x_i^2} \tag{2}$$

- We can show that this is unbiased because we can treat the x_i values as fixed constants, so

$$\mathbb{E}\,\hat{\beta} = \beta + \sum_{i=1}^{n} \left(\frac{x_i}{\sum_{i=1}^{n} x_i^2}\right) \mathbb{E}\,\epsilon_i = \beta \tag{3}$$

Multivariate Regression Model with Fixed Regressors

- This result generalises to the multivariate model with k regressors:

$$Y = X\beta + \epsilon \tag{4}$$

where Y and ϵ are $n \times 1$ vectors, with ϵ containing i.i.d. errors with finite variance, X is an $n \times k$ matrix and β is a $k \times 1$ vector of coefficients.

- We can show that the OLS estimators are unbiased as follows

$$\begin{aligned}
\mathbb{E}\,\hat{\beta} &= \mathbb{E}\left(\left(X'X\right)^{-1}X'Y\right) && (5)\\
&= \mathbb{E}\left(\left(X'X\right)^{-1}X'\left(X\beta+\epsilon\right)\right) && (6)\\
&= \mathbb{E}\left(\beta+\left(X'X\right)^{-1}X'\epsilon\right) && (7)\\
&= \beta+\left(X'X\right)^{-1}X'\,\mathbb{E}\,\epsilon && (8)\\
&= \beta && (9)
\end{aligned}$$

Fixed Regressors and Normal Errors

- Suppose the i.i.d. error term ϵ is normally distributed, $\epsilon \sim N\left(0, \sigma^2 I_n\right)$.
- The formula

$$\hat{\beta} = \beta + \left(X'X\right)^{-1} X'\epsilon \tag{10}$$

  tells us that, when we treat the X as a fixed set of numbers, the OLS estimator is a linear function of a normal variable. So $\hat{\beta}$ is also normally distributed.

- We calculate the covariance of the coefficient estimates as

$$\begin{aligned}
\mathbb{E}\left(\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'\right) &= \mathbb{E}\left[\left(X'X\right)^{-1} X'\epsilon\epsilon' X\left(X'X\right)^{-1}\right] && (11) \\
&= \sigma^2\left[\left(X'X\right)^{-1} X'X\left(X'X\right)^{-1}\right] && (12) \\
&= \sigma^2\left(X'X\right)^{-1} && (13)
\end{aligned}$$

- So $\hat{\beta} \sim N\left(\beta, \sigma^2\left(X'X\right)^{-1}\right)$.

Stochastic Regressors Model

- In reality, the data in the X matrix should usually be thought of as outcomes of random variables, just like the Y variable.
- As long as the X variables are independent of all values in the ϵ matrix, then one can proceed in the same way as though the regressors are fixed.
- For example, one can demonstrate unbiasedness as

$$\begin{aligned} \mathbb{E}\,\hat{\beta} &= \mathbb{E}\left(\beta + \left(X'X\right)^{-1} X'\epsilon\right) && (14)\\ &= \beta + \mathbb{E}\left(\left(X'X\right)^{-1} X'\right)\mathbb{E}\,\epsilon && (15)\\ &= \beta && (16)\end{aligned}$$

- If the errors are normal, then given the realised value of X one can say

$$\hat{\beta}|X \sim N\left(\beta, \sigma^2 \left(X'X\right)^{-1}\right) \quad (17)$$

- This means that, for each possible X, the distribution of β takes on a particular normal value, so the usual t and F statistics also work. (That said, the unconditional distribution of β depends upon the distribution of X and generally will not be Gaussian.)

More on the Independence Assumption

- One might be tempted to think that in the model

$$y_i = \beta x_i + \epsilon_i \quad i = 1, 2...n. \tag{18}$$

that one only needs $\mathbb{E}(x_k \epsilon_k) = 0$ for OLS to be unbiased, i.e. that each error term only needs to be independent of the values of x that occur for the same observation.

- However, one needs a stronger assumption: x_k needs to be independent of all the ϵ_i values for $i = 1, 2..., n$. From equation (2) we can write

$$\mathbb{E}\,\hat{\beta} = \beta + \mathbb{E}\left(\sum_{i=1}^{n}\left(\frac{x_i}{\sum_{i=1}^{n} x_i^2}\right)\epsilon_i\right) \tag{19}$$

- Even if ϵ_j is independent of x_j, it may not be independent of all of the x values that make up the sum $\sum_{i=1}^{n} x_i^2$. This can induce a correlation between $\left(\frac{x_i}{\sum_{i=1}^{n} x_i^2}\right)$ and ϵ_i which leads to OLS being biased.
- We will see that $AR(n)$ time series regressions are an important example of this kind of bias.

Part II

Asymptotic Distributions with Stochastic Regressors

A New Way of Looking at OLS Estimators

- You know the OLS formula in matrix form $\hat{\beta} = (X'X)^{-1} X'Y$. There is a useful way to restate this that allows us to make a clear connection to the WLLN and the CLT.
- Consider the case of a regression with 2 variables and 3 observations. The X matrix is thus

$$X = \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{pmatrix} \tag{20}$$

- This means we can write

$$X'X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix} \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{pmatrix} \tag{21}$$

$$= \begin{pmatrix} x_{11}^2 + x_{12}^2 + x_{13}^2 & x_{11}x_{21} + x_{12}x_{22} + x_{13}x_{23} \\ x_{21}x_{11} + x_{22}x_{12} + x_{23}x_{13} & x_{21}^2 + x_{22}^2 + x_{23}^2 \end{pmatrix} \tag{22}$$

A New Way of Looking at OLS Estimators

- Let x_1 be a column vector containing the 2 datapoints on the explantory variables from the first observation

$$x_1 = \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} \tag{23}$$

- Multplying this vector by its transpose, we get

$$x_1 x_1' = \begin{pmatrix} x_{11}^2 & x_{11}x_{21} \\ x_{21}x_{11} & x_{22}^2 \end{pmatrix} \tag{24}$$

- Going back and comparing this with equation (22) from the previous slide, we get the following new way of describing the $X'X$ matrix

$$X'X = \sum_{i=1}^{n} x_i x_i' \tag{25}$$

A New Formula for the OLS Estimator

- It follows from the previous slides that we can re-write the matrix formula for the OLS estimator as

$$\hat{\beta} = (X'X)^{-1} X'Y = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i y_i' \right) \tag{26}$$

- In the same way, we can write

$$\hat{\beta} = \beta + (X'X)^{-1} X'\epsilon = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i \epsilon_i \right) \tag{27}$$

- This formula is useful because it explains how the OLS estimator depends upon sums of random variables. This allows us to use the Weak Law of Large Numbers and the Central Limit Theorem to establish the limiting distribution of the OLS estimator.

IID Stochastic Regressors

- Consider the case in which ϵ_i are i.i.d. with zero mean and the x variables are i.i.d. each period and there exists a matrix Q_{xx} such that

$$\left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right) \xrightarrow{p} Q_{xx} \quad (28)$$

- This means that

$$\left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \xrightarrow{p} Q_{xx}^{-1} \quad (29)$$

- We can also say that $\mathbb{E}\left(x_i \epsilon_i'\right) = 0$. We can use the multivariate CLT to show that

$$\frac{1}{n} \sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} N\left(0, \Omega\right) \quad (30)$$

where

$$\Omega = \mathbb{E}\left(x_i x_i' \epsilon_i^2\right) \quad (31)$$

Limiting Distribution with IID Stochastic Regressors

- The OLS estimator is

$$\hat{\beta} = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i \epsilon_i \right) \tag{32}$$

with

$$\left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \xrightarrow{p} Q_{xx}^{-1} \tag{33}$$

$$\frac{1}{n} \sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} N(0, \Omega) \tag{34}$$

- Using the Slutsky's Theorem results combining convergence in probability with convergence in distribution, we can say

$$\sqrt{n} \left(\hat{\beta} - \beta \right) \xrightarrow{d} N\left(0, Q_{xx}^{-1} \Omega Q_{xx}^{-1}\right) \tag{35}$$

Unknown Covariance Matrix

- Using the terminology described in the last set of notes, we have shown that

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \frac{1}{n} Q_{xx}^{-1} \Omega Q_{xx}^{-1}\right) \tag{36}$$

- However, we don't know the values of the long-run average Q_{xx} or the covariance matrix Ω. As described in the last notes, though, we can substitute for these with consistent estimators. WLLN tells

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} Q_{xx}^{-1} \tag{37}$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \epsilon_i^2 \xrightarrow{p} \Omega \tag{38}$$

- This means we can substitute sample means for the true population means and base tests on the asymptotic distribution

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{n} x_i x_i' \epsilon_i^2\right) \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\right) \tag{39}$$

Non-Identically Distributed Stochastic Regressors

- In practise, it's unlikely that independent x_i observations will be identically distributed. What if the x_i's are independently but differently distributed?
- One can still prove a Central Limit Theorem for this case. The previous CLT that we proved, for i.i.d. variables, is known as the Lindberg-Levy CLT. Another useful result is the Lindberg-Feller Central Limit Theorem: If a sequence of observations $\{y_1, y_2, ..., y_n\}$ are independently distributed with mean μ and each with a different finite variance σ_i^2 then

$$\sqrt{n}\left(\bar{y}_n - \mu\right) \xrightarrow{d} N\left(0, \bar{\sigma}^2\right). \tag{40}$$

where

$$\bar{\sigma}^2 = \lim_{n\to\infty} \bar{\sigma_n}^2 \tag{41}$$

$$\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 \tag{42}$$

- As usual, there is a natural multivariate extension of this result in which the covariance matrix of the limiting distribution is an average of the covariance matrices of the differently distributed x_is.