Introductory Econometrics

Further Issues: Large Sample Properties of OLS

Monash Econometrics and Business Statistics

Semester 2, 2018

Recap

- We studied how to develop autoregressive distributed lag (ARDL) models that allow for a dynamic relationship between the dependent and independent variables
- We learnt how the parameters of an ARDL model can measure the immediate and the long-run effect of the independent variables on the dependent variable.
- We learnt that with time series we can even have univariate models, i.e. models that do not have any independent variables and only use the history of a time series to predict its future
- From univariate time series models, we studied the autoregressive (AR) models only
- We saw that the AR model can produce time series that are stationary, time series that wander about, and time series that explode

Recap

- We studied the concept of mean reversion and stationarity more formally
- We made the distinction between a stationary and non-stationary time series
- A stationary time series has:
  1. constant mean over time
  2. constant variance over time
  3. the covariance between any two observations depends only on the time interval separating them and not on time itself.
- If one of these conditions are violated then the time series is said to be non-stationary.
- We learnt that if the variables are stationary and the errors are not predictable from the past, then the OLS estimator is not unbiased, but it is a consistent estimator, and we can use the usual t-test and F-test in large samples

Lecture Outline

- Asymptotic properties of the OLS estimator:

1. Consistency: What happens if all we have is that errors are uncorrelated with regressors, and we do not have $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$? OLS will not be unbiased, but will be *consistent* (textbook reference 5-1)
2. Asymptotic Normality: What happens if errors are not normally distributed? The distribution of the OLS estimator will be approximately normal in large samples (textbook reference 5-2)

- As we have seen, in many business and economics examples we have time series data (monthly production and sales figures, quarterly Australian CPI data, a family's annual income and expenditure figures, hourly electricity load in Melbourne, minute by minute stock prices)
- Dynamic models are of the form:

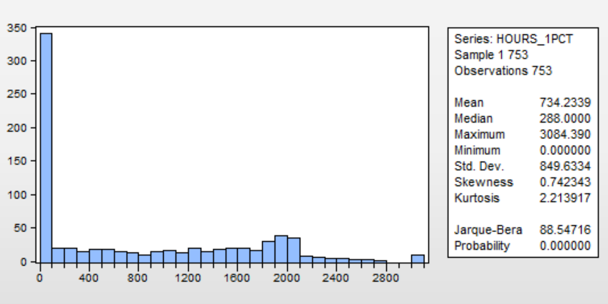
$$food_t = \beta_0 + \beta_1 food_{t-1} + \beta_2 inc_t + u_t,\ \ t = 2, \ldots, n$$

- Note that $t = 2, \ldots, n$ because on the right hand side we have $food_{t-1}$
- Sample is no longer random and therefore

$$E(\mathbf{u} \mid \mathbf{X}) = E\left(\begin{pmatrix} u_2 \\ u_3 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} \middle| \begin{pmatrix} 1 & food_1 & inc_2 \\ 1 & food_2 & inc_3 \\ \vdots & \vdots & \vdots \\ 1 & food_{n-2} & inc_{n-1} \\ 1 & food_{n-1} & inc_n \end{pmatrix} \right) \neq \mathbf{0}$$

- Also, in many business and economics examples, variables are not normally distributed, for example

Married women annual hours worked



- Is the OLS machinery practically useless in such situations?

- Two amazing mathematical results:

1. The Law of Large Numbers: Sample averages converge to population means as $n \to \infty$
2. The Central Limit Theorem: The distribution of sample sums or averages becomes closer and closer to normal as n gets large

- These results make the machinery that we have developed so far relevant to the real world.
- Because of these results, we can use OLS to estimate dynamic models in which $E(\mathbf{u} \mid \mathbf{X}) \neq \mathbf{0}$ as long as for each t, u_t is uncorrelated with $x_{t1}, x_{t2}, \ldots, x_{tk}$ and the past history, and we can make inference based on the OLS estimator based on t and F tests even if the errors are not normally distributed (the audience goes WOW!)

Law of Large Numbers and Consitency of the OLS

- Consider a single variable y with $E(y) = \mu$ and $Var(y) = \sigma^2$
- Suppose we have a random sample of n observations from this population
- Denote the sample average by $\bar{y}$
- We know $E(\bar{y}) = \mu$ and $Var(\bar{y}) = \sigma^2/n$
- As $n \to \infty$, $Var(\bar{y}) \to 0$, and the chance of $\bar{y}$ being anything other than μ goes to zero.
- We say that $\bar{y}$ converges in probability to μ and we write that $\text{plim}(\bar{y}) = \mu$ or $\bar{y} \xrightarrow{p} \mu$. This is the Law of Large Numbers
- If an estimator converges in probability to the population parameter that it estimates, we say that the estimator is *consistent*
- The sample mean is a consistent estimator of the population mean

- Unlike the expected value, which could only be applied to linear combination of random variables, plim can go through non-linear combinations and smooth functions as well
- For example,

$$
\begin{aligned}
E(\bar{y}^2) &\neq \mu^2, \text{ but } \operatorname{plim}(\bar{y}^2) = \mu^2 \\
E\left(\frac{1}{\bar{y}}\right) &= ?, \text{ but } \operatorname{plim}\left(\frac{1}{\bar{y}}\right) = \frac{1}{\mu} \text{ provided } \mu \neq 0 \\
E\left(\frac{1}{\hat{\sigma}_y^2}\right) &= ?, \text{ but } \operatorname{plim}\left(\frac{1}{\hat{\sigma}_y^2}\right) = \frac{1}{Var(y)}
\end{aligned}
$$

- Why is the Law of Large Numbers important for the OLS estimator?
- Recall:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

- Consider $\mathbf{X}'\mathbf{u}$

$$\begin{aligned}
\mathbf{X}'\mathbf{u} &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & \vdots & & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n} u_i \\ \sum_{i=1}^{n} (x_{i1} u_i) \\ \vdots \\ \sum_{i=1}^{n} (x_{ik} u_i) \end{pmatrix}
\end{aligned}$$

- This implies that:

$$\frac{1}{n}\mathbf{X}'\mathbf{u} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} u_i \\ \frac{1}{n}\sum_{i=1}^{n}(x_{i1}u_i) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n}(x_{ik}u_i) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} E(u) \\ E(x_1 u) \\ \vdots \\ E(x_k u) \end{pmatrix}$$

- If we have $E(u) = 0, E(x_1 u) = 0, \ldots, E(x_k u) = 0$, then $\text{plim}(\frac{1}{n}\mathbf{X}'\mathbf{u}) = \mathbf{0}$
- This is of course satisfied when $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$. However, $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$ is more than necessary and do not hold when we have time series data and dynamic models. We only need that this period's error to be uncorrelated with previous periods' information, which is reasonable

- We can also show that:

$$
\begin{aligned}
\frac{1}{n}\mathbf{X}'\mathbf{X} &= \begin{pmatrix} 1 & \bar{x}_1 & \cdots & \bar{x}_k \\ \bar{x}_1 & \frac{1}{n}\sum_{i=1}^{n}(x_{i1}^2) & \cdots & \frac{1}{n}\sum_{i=1}^{n}(x_{i1}x_{ik}) \\ \vdots & \vdots & & \vdots \\ \bar{x}_k & \frac{1}{n}\sum_{i=1}^{n}(x_{ik}x_{i1}) & \cdots & \frac{1}{n}\sum_{i=1}^{n}(x_{ik}^2) \end{pmatrix} \\
&\xrightarrow{p} \quad \mathbf{Q} \text{ an invertible matrix}
\end{aligned}
$$

- Hence we have

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
&= \boldsymbol{\beta} + (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}'\mathbf{u}) \\
\text{plim}(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + \text{plim}(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}\,\text{plim}(\frac{1}{n}\mathbf{X}'\mathbf{u}) \\
&= \boldsymbol{\beta} + \mathbf{Q}^{-1}\mathbf{0} = \boldsymbol{\beta}
\end{aligned}
$$

- Therefore the OLS estimator is consistent

Consistency of the OLS estimator

- Conclusion: Even if we have non-random samples, i.e. $E(\mathbf{u} \mid \mathbf{X}) \neq \mathbf{0}$, as long as $E(u_i) = 0$ and u_i is uncorrelated with x_{i1} to x_{ik}, then the OLS estimator is consistent
- This allows us to use OLS estimator to estimate dynamic models and know that as long as we have a reasonably large sample, our estimates are unlikely to be far from the true parameters

Central Limit Theorem

- The sum (or the average) of n normal random variables (r.v.'s) is normal. However,
- the sum (or the average) of n uniform r.v.'s is not uniform
- the sum (or the average) of n F r.v.'s is not F
- the sum (or the average) of n Bernouli r.v.'s (1 with probability p and 0 with probability $1 - p$) is not Bernouli
- ...

- However, if n is large,
  - the sum (or the average) of n uniform r.v.'s is approximately normal;
  - the sum (or the average) of n F r.v.'s is approximately normal;
  - and even the sum (or the average) of n Bernouli r.v.'s (1 with probability p and 0 with probability $1 - p$) is approximately normal!
  - In fact, the sum (or the average) of n r.v.'s from any distribution with a finite variance is approximately normal!
- See onlinestatbook.com/stat_sim/sampling_dist/. Notice that we do not need very many observations to get a good approximation.

- This means that for a sample of n observations of any r.v. y with mean μ and variance σ^2 with any arbitrary distribution, we can say:

$$\begin{aligned} \sum_{i=1}^{n} y_i &\overset{a}{\sim} N(n\mu, n\sigma^2) \\ \bar{y} &\overset{a}{\sim} N(\mu, \sigma^2/n) \end{aligned}$$

- Either one implies:

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0, 1)$$

- In fact as $n \to \infty$ this last one becomes exact, and not approximate:

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

- This last result is the Central Limit Theorem

- Why is the Central Limit Theorem important for the OLS estimator?
- Recall:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = \boldsymbol{\beta} + (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}'\mathbf{u})$$

- This shows that $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is a linear combination of averages, and since each of those averages will be approximately normal in large samples, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ will be approximately normal
- In fact as $n \to \infty$,

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \xrightarrow{d} N(0,1)$$

- Therefore, we can use t and F tests to do inference (recall that t with a large degree of freedom is approximately a $N(0,1)$)

Asymptotic normality of the OLS estimator

- Conclusion: For any error distribution, as long as the sample size is large, the OLS estimator is approximately normal, and we can base our statistical inference on the usual t and F tests
- This will allow us to use OLS even if the distribution of the dependent variable is far from normal, and may even have spikes on some values, such as the spike at zero for the married women's work hours.

Summary

- Let's put all of these together:
  - In $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$, $i = 1, \ldots, n$
  - if $E(u_i) = 0$ and u_i is uncorrelated with x_{i1} to x_{ik}
  - and if x variables are linearly independent
  - and if $Eu_i^2 = \sigma^2$ for all i
  - then the OLS estimator is consistent and asymptotically normal

$$\widehat{\boldsymbol{\beta}} \stackrel{a}{\sim} N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

- If we have HTSK, we use robust standard errors, or we use appropriate weights and use weighted least squares
- These properties allow us to use the OLS estimator to estimate models based on time series data as well as cross section data, and they also show that we can do valid inference using OLS based on data from any distribution, as long as we have a large sample