Introductory Econometrics

Large Sample Properties of OLS

Monash Econometrics and Business Statistics

2022

Recap

The multiple regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i, \;\; i = 1, 2, ...n.$$

- A1 model is linear in parameters: $y = X\beta + u$.
- A2 columns of X are linearly independent.
- A3 conditional mean of errors is zero: $E(u|X) = 0$.
- A4 homoskedasticity and no serial correlation: $Var(u|X) = \sigma^2 I_n$.
- A5 errors are normally distributed: $u|X \sim N(0, \sigma^2 I_n)$.

Recap

- We studied time series models as AR and ARDL models.
- AR models only use the history of a time series to predict its future.
- ARDL models measure immediate and long-run effects.
- We made the distinction between (non-) stationary time series.
- If the variables are stationary and the errors white noise:

The OLS estimator of the parameters of a dynamic model is reliable and we can use t and F tests provided that the sample size is large.

Why?

Zero conditional mean

The linear regression model with time series data

$$y_t = \beta_0 + \beta_1 x_t + u_t, \;\; t = 1, 2, ...n.$$

A3 conditional mean of errors is zero: $E(u|X) = 0$.

Which implies that $corr(u_t, x_1) = \cdots = corr(u_t, x_n) = 0$ for $t = 1, 2, ...n$.

Violation A3 with time series

The AR(1) model

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t,\ \ t = 1, 2, ...n.$$

- It follows that $y_{t-1} = \beta_0 + \beta_1 y_{t-2} + u_{t-1}$.
- Hence $corr(u_{t-1}, y_{t-1}) \neq 0$.
- This violates A3: $corr(u_{t-1}, y_{t-1}) = 0$.

Violation A3 with time series

In general, the OLS estimator is not unbiased with time series.

So what do we do?

- We show that the OLS estimator will be consistent.
- We show that the distribution of the OLS estimator will be approximately normal in large samples.

Lecture Outline

- Consistency
- Asymptotic normality
- Homoskedasticity and serial correlation

Consistency

Consider a set of i.i.d. random variables $(X_1, \ldots, X_n)$ with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ for $i = 1, \ldots, n$.

- Define $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.
- Recall that $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \sigma^2/n$.
- $\bar{X}$ is an unbiased estimator of the parameter μ.

Consistency

The sample mean is also a consistent estimator of μ:

- As $n \to \infty$, $Var(\bar{y}) \to 0$.
- The chance of $\bar{y}$ being anything other than μ goes to zero.
- We say that $\bar{y}$ converges in probability to μ.
- We write that $\text{plim}(\bar{y}) = \mu$ or $\bar{y} \xrightarrow{p} \mu$.

If an estimator converges in probability to the population parameter that it estimates, we say that the estimator is consistent.

Consistency

A consistent estimator can be biased:

- Define $\tilde{X} = \bar{X} + \frac{1}{n}$.
- It holds that $E(\tilde{X}) = \mu + \frac{1}{n}$ and $Var(\bar{X}) = \sigma^2/n$.
- As $n \to \infty$, $E(\tilde{y}) \to \mu$ and $Var(\tilde{y}) \to 0$.

Since the estimator converges in probability to the population parameter that it estimates, we say that the estimator is consistent.

Consistency

The multiple regression model

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t, \;\; t = 1, 2, ...n.$$

A1 model is linear in parameters: $y = X\beta + u$.

A2 columns of X are linearly independent.

A3 conditional mean of errors is zero: $E(u_t|x_t) = 0$ (NEW).

Under these assumptions, the OLS estimator is consistent: $\text{plim}(\hat{\beta}) = \beta$.

Consistency

The NEW assumption does not always hold:

A3 conditional mean of errors is zero: $E(u_t|x_t) = 0$ (NEW).

For instance, consider

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t, \quad |\theta_1| < 1,$$
$$u_t = \theta_i u_{t-1} + e_t, \qquad e_t \sim i.i.d.(0, \sigma^2).$$

- Since $y_{t-1} = \beta_0 + \beta_1 y_{t-2} + u_{t-1}$,
- we have $Corr(y_{t-1}, u_{t-1})$ and $Corr(u_{t-1}, u_t)$
- and $Corr(u_t, y_{t-1})$

Asymptotic normality

The multiple regression model with time series

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t, \;\; t = 1, 2, ...n.$$

A1 model is linear in parameters: $y = X\beta + u$.

A2 columns of X are linearly independent.

A3 conditional mean of errors is zero: $E(u_t|x_t) = 0$ (NEW).

A4 homoskedasticity and no serial correlation:
$Var(u_t|x_t) = \sigma^2$ for all t and $E(u_t u_s|x_t, x_s) = 0$ for all $t \neq s$ (NEW).

Under these assumptions, the OLS estimator is asymptotically normal:

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

Asymptotic normality

Note that we do not require an assumption on the error distribution:

A5 errors are normally distributed: $u|X \sim N(0, \sigma^2 I_n)$.

This follows from the Central limit theorem:

- average of n random variables from any distribution with a finite variance is approximately normal if n is large.
- See onlinestatbook.com/stat_sim/sampling_dist/.
- $\widehat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{y} = \boldsymbol{\beta} + (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{u} = \boldsymbol{\beta} + (\frac{1}{n}\mathrm{X}'\mathrm{X})^{-1}(\frac{1}{n}\mathrm{X}'\mathrm{u})$.
- This shows that $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is a linear combination of averages.
- $\widehat{\boldsymbol{\beta}} \overset{a}{\sim} N(\boldsymbol{\beta}, \sigma^2(\mathrm{X}'\mathrm{X})^{-1})$.

Asymptotic normality

- For any error distribution, as long as the sample size is large, the OLS estimator is approximately normal.
- As $n \to \infty$,
$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \xrightarrow{d} N(0,1),$$
since t with a large degree of freedom is approximately a $N(0,1)$.
- 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.

Homoskedasticity and serial correlation

In case this assumption does not hold:

A4 homoskedasticity and no serial correlation:
$Var(u_t|x_t) = \sigma^2$ for all t and $E(u_t u_s|x_t, x_s) = 0$ for all $t \neq s$ (NEW).

We use OLS with HAC standard errors.

Since with time series this assumption rarely holds:

A4 homoskedasticity and no serial correlation: $Var(u|X) = \sigma^2 I_n$.

Even when the OLS estimator in a time series model is unbiased,
it is rarely the most efficient unbiased estimator.

Summary

The multiple regression model with time series

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t, \;\; t = 1, 2, ...n.$$

A1 model is linear in parameters: $y = X\beta + u$.

A2 columns of X are linearly independent.

A3 conditional mean of errors is zero: $E(u_t|x_t) = 0$ (NEW).

A4 homoskedasticity and no serial correlation:
$Var(u_t|x_t) = \sigma^2$ for all t and $E(u_t u_s|x_t, x_s) = 0$ for all $t \neq s$ (NEW).

Under these assumptions, the OLS estimator is asymptotically normal:

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

- If A4 does not hold, use OLS with HAC standard errors.
- These assumptions allow us to use the OLS estimator to estimate models based on time series data as well as cross section data, as long as we have a large sample.
- 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.