ECON 61001: *Lecture 4*

Alastair R. Hall

The University of Manchester

Outline of this lecture

- Stochastic regressors
- Large sample theory: concepts & Limit Theorems

- Large sample behaviour of OLS
- Large sample inference in models estimated from cross-section data

Stochastic regressors

Recall that so far our model has been:

$$y = X\beta_0 + u$$

where

- *CA1*: true model is: $y = X\beta_0 + u$.
- *CA2*: X is *fixed in repeated samples*.
- *CA3*: X is rank k.
- *CA4*: $E[u] = 0$.
- *CA5*: $Var[u] = \sigma_0^2 I_T$.
- *CA6*: $u \sim$ Normal.

Now consider model with stochastic regressors.

Assumptions:

- *SR1*: true model is: $y = X\beta_0 + u$.
- *SR2*: X is stochastic.
- *SR3*: X is rank k with probability 1.
- *SR4*: $E[u|X] = 0$.
- *SR5*: $Var[u|X] = \sigma_0^2 I_T$.
- *SR6*: $u|X \sim$ Normal.

$\Rightarrow y|X \sim N(X\beta_0, \sigma_0^2 I_T)$.

Stochastic regressor model

Results discussed in Lecture 2 before still go through although with some additional conditions: *e.g.*

$$E[\hat{\beta}_T] = \beta_0 + E_X\left[E_{u|X}\left[(X'X)^{-1}X'u\right]\right]$$

and provided the expectation on the rhs exists then,

$$E[\hat{\beta}_T] = \beta_0 + E_X\left[(X'X)^{-1}X'E_{u|X}[u]\right] = \beta_0, \quad \text{using } SR4.$$

Distributions of test statistics also goes through but depend crucially on *SR6*. If distribution is not normal then inference methods discussed in previous lectures are invalid. $\rightarrow$ alternative approach using large sample theory.

Recall that

$$\hat{\beta}_T = (X'X)^{-1}X'y = \beta_0 + (X'X)^{-1}X'u$$

$$= \beta_0 + \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t u_t$$

So $\{\hat{\beta}_T; T = k, k+1, \ldots\}$ is a stochastic sequence indexed by T.

Large sample theory rests on considering how statistics of interest ($\hat{\beta}_T$, t-stats *etc*) behave as $T \to \infty$.

Our large sample analysis rests on two key concepts: "convergence in probability" and "convergence in distribution".

Consider stochastic sequence $\{V_T;\ T = 1, 2, \ldots\}$ and random variable V.

V_T is said to converge in probability to V if $P(|V_T - V| < \epsilon) \to 1$ for any $\epsilon > 0$ as $T \to \infty$, and is denoted by $V_T \xrightarrow{p} V$.

This definition holds for rv V. If V is a *degenerate* rv (that is, a constant) then we have two important items of terminology.

Definition: If $V_T \xrightarrow{p} c$ where c is a constant then c is referred to as the probability limit of V_T and this is written as $plim V_T = c$.

Definition: If $\hat{\theta}_T$ is an estimator of the unknown parameter θ_0 and $\hat{\theta}_T \xrightarrow{p} \theta_0$ then $\hat{\theta}_T$ is said to be a consistent estimator of θ_0.

Convergence in probability

Let $\{M_T;\ T = 1, 2, \ldots\}$ be a sequence of random matrices and M be random matrix (M_T, M are $p \times q$).

Then

$$M_T \xrightarrow{p} M \text{ iff } M_{T,i,j} \xrightarrow{p} M_{i,j} \text{ for } i = 1, 2, \ldots p,\ j = 1, 2, \ldots q,$$

where i, j subscript denotes $i - j^{th}$ element of the matrix in question.

Slutsky's theorem: Let $\{V_T\}$ be a sequence of $r \times 1$ random vectors (or matrices) which converge in probability to the random vector (or matrix) V and let $f(.)$ be a real- valued vector of continuous functions then $f(V_T) \xrightarrow{p} f(V)$.

Convergence in distribution

The sequence of random variables $\{V_T\}$ with corresponding distribution functions $\{F_T(\cdot)\}$ converges in distribution to the random variable V with distribution function $F(X)$ if and only if $F_T(c) \to F(c)$ as $T \to \infty$ at all points of continuity $\{c\}$ of $F(.)$.

- Convergence in distribution is denoted by $V_T \xrightarrow{d} V$.

- The distribution of V is known as the *limiting distribution* of V_T.

Note: If $V_T \xrightarrow{p} V$ then $V_T \xrightarrow{d} V$ but reverse implication does not hold unless $V = c$, a constant.

Recall that

$$\hat{\beta}_T - \beta_0 = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t u_t$$

So large sample behaviour depends on:

- a random matrix $\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}$,
- a random vector $\sum_{t=1}^{T} x_t u_t$.

Their large sample behaviour is specified by two fundamental Limit Theorems:

- the Weak Law of Large Numbers (WLLN)
- the Central Limit Theorem (CLT)

Let $\{v_t,\ t = 1, 2, \ldots T\}$ be random vectors with $E[v_t] = \mu_t$.

Weak Law of Large Numbers: Subject to certain conditions,

$$T^{-1}\sum_{t=1}^{T}(v_t - \mu_t) \xrightarrow{p} 0.$$

Central Limit Theorem: Subject to certain conditions,

$$T^{-1/2}\sum_{t=1}^{T}(v_t - \mu_t) \xrightarrow{d} N(0, \Omega).$$

where

$$\Omega \;=\; lim_{T\to\infty} Var\left[\, T^{-1/2}\sum_{t=1}^{T}(v_t \;-\; \mu_t)\right]$$

is a finite positive definite matrix of constants.

In addition, we have the following two useful results:

Large sample behaviour of matrix-vector products:

If $b_T = M_T m_T$ where M_T is a $q \times r$ random matrix and m_T is a $r \times 1$ random vector and:

- $M_T \xrightarrow{p} M$, a matrix of finite constants,
- $m_T \xrightarrow{d} N(0, \Omega)$, Ω is finite p.d. matrix of constants,
- $rank(M\Omega M') = q$,

Then

$$b_T \xrightarrow{d} N(0, M\Omega M')$$

Large sample behaviour of quadratic forms:

If $a_T = m_T' \hat{\Omega}_T^{-1} m_T$ where m_T is a $r \times 1$ random vector and:

- $m_T \xrightarrow{d} N(0, \Omega)$, where Ω is a finite p.d. matrix of constants,
- $\hat{\Omega}_T \xrightarrow{p} \Omega$,

Then

$$a_T \xrightarrow{d} \chi^2_r$$

To develop a large sample analysis for OLS with cross-section data we impose the following assumptions.

- **Assumption CS1**:
  $y_i = x_i'\beta_0 + u_i = \beta_{0,1} + x_{2,i}'\beta_{0,2} + u_i, \qquad i = 1, 2, \ldots, N$

- **Assumption CS2**: $\{(u_i, x_i'), \; i = 1, 2, \ldots N\}$ forms an independent and identically distributed sequence.

- **Assumption CS3**: $E[x_i x_i'] = Q$, finite, p.d.

- **Assumption CS4**: $E[u_i | x_i] = 0$.

- **Assumption CS5**: $Var[u_i | x_i] = \sigma_0^2$, positive, finite constant.

Large sample analysis of OLS

In our notation here, the OLS estimator is

$$\hat{\beta}_N = \left(\sum_{i=1}^{N} x_i x_i' \right)^{-1} \sum_{i=1}^{N} x_i y_i$$

and so using *CS1*, we have

$$\hat{\beta}_N = \beta_0 + \left(\sum_{i=1}^{N} x_i x_i' \right)^{-1} \sum_{i=1}^{N} x_i u_i$$

$$= \beta_0 + \left(N^{-1} \sum_{i=1}^{N} x_i x_i' \right)^{-1} N^{-1} \sum_{i=1}^{N} x_i u_i$$

Under Assumptions *CS1*-*CS5*, we can use the WLLN to deduce:

- $N^{-1} \sum_{i=1}^{N} x_i x_i' \xrightarrow{p} Q$
- $N^{-1} \sum_{i=1}^{N} x_i u_i \xrightarrow{p} E[x_i u_i] = E\left[x_i \, E[u_i \mid x_i] \right] = 0$ (via LIE)

And so using Slutsky's Theorem,

$$
\begin{aligned}
\hat{\beta}_N - \beta_0 &= \left(N^{-1} \sum_{i=1}^{N} x_i x_i' \right)^{-1} N^{-1} \sum_{i=1}^{N} x_i u_i \\
&\xrightarrow{p} Q^{-1} \times 0 = 0
\end{aligned}
$$

$\Rightarrow \hat{\beta}_N \xrightarrow{p} \beta_0$.

So $\hat{\beta}_N$ is consistent for β_0.

Under these conditions, we can also apply the CLT to deduce:

$$N^{-1/2}\sum_{i=1}^{N} x_i u_i \xrightarrow{d} N(0, \Omega)$$

To deduce form of Ω, define $\Omega_N = Var[N^{-1/2}\sum_{i=1}^{N} x_i u_i]$.

Using $(u_i, x_i) \sim i.i.d.$ and $E[x_i u_i] = 0$, we have

$$\begin{aligned}\Omega_N &= N^{-1}\sum_{i=1}^{N} Var[x_i u_i] = N^{-1}\sum_{i=1}^{N} E[u_i^2 x_i x_i'] \\ &= N^{-1}\sum_{i=1}^{N} E\left[\, E[u_i^2 | x_i] x_i x_i' \,\right] = \sigma_0^2 Q\end{aligned}$$

So $\Omega = lim_{N\to\infty}\Omega_N = \sigma_0^2 Q$.

As we have

$$N^{1/2}(\hat{\beta}_N - \beta_0) = \left(N^{-1}\sum_{i=1}^{N} x_i x_i'\right)^{-1} N^{-1/2}\sum_{i=1}^{N} x_i u_i$$

and:

- $N^{-1}\sum_{i=1}^{N} x_i x_i' \xrightarrow{p} Q$, nonsingular
- $N^{-1/2}\sum_{i=1}^{N} x_i u_i \xrightarrow{d} N(0, \sigma_0^2 Q)$.

it follows that

$$N^{1/2}(\hat{\beta}_N - \beta_0) \xrightarrow{d} N\left(0, \sigma_0^2 Q^{-1}\right)$$

To perform inference, we need to estimate variance.

Use

$$\hat{\Omega}_N = \hat{\sigma}_N^2 N^{-1} X'X$$

So, for example, an approximate $100(1-\alpha)\%$ confidence interval for $\beta_{0,\ell}$ based on limiting distribution is:

$$\hat{\beta}_{N,\ell} \pm z_{1-\alpha/2}\hat{\sigma}_N\sqrt{m_{\ell,\ell}}$$

where

- $m_{\ell,\ell}$ is the ℓ^{th} main diagonal element of $(X'X)^{-1}$.
- $z_{1-\alpha/2}$ is $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution

Suppose we wish to test $H_0 : R\beta_0 = r$ vs $H_1 : R\beta_0 \neq r$.

Test statistic

$$W_N = N(R\hat{\beta}_N - r)' \left[R(N^{-1}X'X)^{-1}R' \right]^{-1} (R\hat{\beta}_N - r)/\hat{\sigma}^2_N$$

Then under H_0: $W_N \xrightarrow{d} \chi^2_{n_r}$.

(Note: $W_N = n_r F$ from Lecture 3.)

Suppose we wish to test: $H_0 : g(\beta_0) = 0$ vs $H_1 : g(\beta_0) \neq 0$.
Assume:

- $g(\cdot)$ is a $n_g \times 1$ vector of continuous differentiable functions
- $G(\bar{\beta}) = \partial g(\beta)/\partial \beta'|_{\beta=\bar{\beta}}$ with $rank\{G(\beta_0)\} = n_g$

Can test H_0 using:

$$W_N^{(g)} = N g(\hat{\beta}_N)' \left[G(\hat{\beta}_N)(N^{-1}X'X)^{-1}G(\hat{\beta}_N)' \right]^{-1} g(\hat{\beta}_N)/\hat{\sigma}_N^2$$

and under H_0: $W_N^{(g)} \xrightarrow{d} \chi^2_{n_g}$.

Proof based on so-called "Delta method".

Further reading

- Notes: Sections 2.11, 3.1-3.2
- Greene:
  - convergence in probability, Section D.2.1 p.1107-1110; Slutsky's Theorem D.2.3, p.1113-1114
  - convergence in distribution, Section D.2.5, p.1116-1118.
  - WLLN, Section D.2.2, p.1110-1113.
  - CLT, Section D.2.6, p.1118-1123.
  - Large sample behaviour of OLS, Section 4.4.1-4.4.3, p.103-108.