Advanced Econometrics

Chapter 5: Large-sample properties of the LSE

In Choi

Sogang University

Stochastic convergence

- Suppose that $\{X_n\}$ is a sequence of random variables with a corresponding sequence of distribution functions $\{F_n\}$.

Convergence in distribution If $F_n(x) \longrightarrow F(x)$ at every continuity point x of F, F_n is said to converge weakly to F, written $F_n \Rightarrow F$. In this case, $\{X_n\}$ is said to converge in distribution to X where X is a random variable with distribution function F, written $X_n \xrightarrow{d} X$.

Example

Let Z_j be a sequence of iid random variables having a binomial distribution $B(1, \frac{1}{2})$. By the central limit theorem

$$X_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(Z_j - \frac{1}{2} \right) \xrightarrow{d} N(0, \frac{1}{4})$$

as $n \longrightarrow \infty$.

Stochastic convergence

Convergence in probability If X is a random variable, and for all $\varepsilon > 0$

$$\lim_{n \to \infty} P\left(|X_n - X| > \varepsilon\right) = 0,$$

X_n is said to converge in probability to X, written $X_n \xrightarrow{p} X$. X is known as the probability limit of X_n, written $X = \text{plim} X_n$.

Example

Let Z_j be a sequence of iid random variables having a binomial distribution $B(1, \frac{1}{2})$. Let $Y_n = \frac{1}{n}\sum_{j=1}^{n}\left(Z_j - \frac{1}{2}\right)$Then, by the Chebyshev's inequality

$$P\left(|Y_n| > \varepsilon\right) \leq \frac{E(Y_n^2)}{\varepsilon^2} = \frac{1}{\varepsilon^2 n^2} \times \frac{n}{4} = \frac{1}{4n} \to 0 \text{ as } n \to \infty.$$

Thus, $Y_n \xrightarrow{p} 0$

Stochastic convergence

Convergence in mean square If

$$\lim_{n\to\infty} E\left(X_n - X\right)^2 = 0,$$

X_n is said to converge in mean square to X, written $X_n \stackrel{m.s.}{\to} X$.

Example

In the previous example, we showed $Y_n \stackrel{m.s.}{\to} 0$.

Some useful results regarding stochastic convergence are:

1. $X_n \xrightarrow{p} X$ and $g(\cdot)$ is a continuous function

$$\Rightarrow g(X_n) \xrightarrow{p} g(X).$$

Example

Let

$$X_n = \begin{cases} 1 & \text{with probability } \frac{1}{n} \\ 0 & \text{with probability } 1 - \frac{1}{n} \end{cases}.$$

Obviously, $X_n \xrightarrow{p} 0$. Let $g(x) = x + 1$. Then, $g(X_n) \xrightarrow{p} g(0) = 1$.

Stochastic convergence

2. Suppose that $Y_n \xrightarrow{d} Y$ and $X_n \xrightarrow{p} c$ (a constant). Then

$$X_n + Y_n \xrightarrow{d} c + Y$$

$$X_n Y_n \xrightarrow{d} cY$$

$$\frac{Y_n}{X_n} \xrightarrow{d} \frac{Y}{c} \text{ when } c \neq 0.$$

3. $X_n \xrightarrow{d} X$ and $g(\cdot)$ is continuous

$$\Rightarrow g(X_n) \xrightarrow{d} g(X).$$

(This is called the continuous mapping theorem)

Example

If $X_n \xrightarrow{d} N(0,1)$, $X_n^2 \xrightarrow{d} \chi^2(1)$.

Stochastic convergence

4. $X_n - Y_n \xrightarrow{p} 0$ and $X_n \xrightarrow{d} X$.

$$\Rightarrow Y_n \xrightarrow{d} X.$$

5. $X_n \xrightarrow{p} X$ implies $X_n \xrightarrow{d} X$. (The converse is not necessarily true.)
6. $X_n \xrightarrow{d} c$ (a constant)

$$\Rightarrow X_n \xrightarrow{p} c.$$

7. $X_n \xrightarrow{m.s.} X$

$$\Rightarrow X_n \xrightarrow{p} X.$$

Stochastic convergence

- If for any $\varepsilon > 0$, there exists $B_\varepsilon < \infty$ such that

$$P\left(\frac{|X_n|}{n^r} > B_\varepsilon\right) < \varepsilon$$

  for all $n \geq 1$, write $X_n = O_p\left(n^r\right)$. $\left(\frac{X_n}{n^r} \text{ is stochastically bounded}\right)$
- If $\text{plim}\frac{X_n}{n^r} = 0$, write $X_n = o_p\left(n^r\right)$.

1. Let $\{X_i, i \geq 1\}$ be a sequence of i.i.d. r.v.s with

$$EX_1 < \infty.$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} EX_1 \text{ as } n \to \infty.$$

2. Let $\{X_i, i \geq 1\}$ be a sequence of independent r.v.s with $EX_i = \mu_i$ and $\frac{1}{n}\sum_{i=1}^{n} \mu_i \to m$. If $E\left|X_i\right|^{1+\delta} \leq B < \infty$ $(\delta > 0)$ for all i, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} m \text{ as } n \to \infty.$$

3. Let $\{X_t, t \geq 1\}$ be a sequence of dependent r.v.s with $EX_t = \mu$. If $Var(X_t) \leq B < \infty$ $(\delta > 0)$ for all t and, $\sum_{j=1}^{n} j^{-1} B_j < \infty$ for all n where $B_j = \sup_t |Cov(X_t, X_{t-j})|$, then

$$\frac{1}{n} \sum_{t=1}^{n} X_t \xrightarrow{p} \mu \text{ as } n \to \infty. \quad (1)$$

Example

Let $X_t = e_t + 0.5 e_{t-1}$ with $e_t \sim iid(0,1)$. Then, $\frac{1}{n} \sum_{t=1}^{n} X_t \xrightarrow{p} 0$ as $n \to \infty$.

Example

Let $X_t = \alpha X_{t-1} + e_t$, $|\alpha| < 1$, $e_t \sim iid(0,1)$. Then, $\frac{1}{n} \sum_{t=1}^{n} X_t \xrightarrow{p} 0$ as $n \to \infty$. (Note: $Cov(X_t, X_{t-j}) = \frac{\alpha^j}{1-\alpha^2}$)

Proof.

Due to Chebyshev's in equality, it suffices to show that $\frac{1}{n^2}Var(S_n) \longrightarrow 0$ as $n \longrightarrow \infty$, where $S_n = \sum_{t=1}^{n} X_t$. Let $Cov(X_t, X_{t-m}) = \sigma_{t,t-m}$. Then,

$$\begin{aligned}
\frac{1}{n^2}Var(S_n) &= \frac{1}{n^2}\sum_{t=1}^{n}\sigma_{t,t} + \frac{2}{n^2}\sum_{t=2}^{n}\sigma_{t,t-1} + \frac{2}{n^2}\sum_{t=3}^{n}\sigma_{t,t-2} + \ldots + \frac{1}{n^2}\sigma_{n,1} \\
&\leq \frac{nB}{n^2} + \frac{2}{n^2}\sum_{m=1}^{n-1}\sum_{t=m+1}^{n}|\sigma_{t,t-m}| \\
&\leq \frac{B}{n} + \frac{2}{n^2}\sum_{m=1}^{n-1}(n-m)B_m = \frac{B}{n} + \frac{2}{n}\sum_{m=1}^{n-1}B_m - \frac{2}{n^2}\sum_{m=1}^{n-1}mB_m.
\end{aligned}$$

The second and third terms converge to zero due to Kronecker's lemma. □

Chebyshev's inequality

$$P(|X| > \varepsilon) \leq \frac{1}{\varepsilon^2} E\left(X^2\right).$$

Proof. Let $Z = 0$ if $|X| \leq \varepsilon$ and $Z = \varepsilon^2$ if $|X| > \varepsilon$. Then, $Z \leq X^2$ and $E(Z) \leq E(X^2)$. But $E(Z) = \varepsilon^2 P(|X| > \varepsilon)$.

Kronecker's lemma $\{a_t\}, \{x_t\}$: sequences of positive real numbers
$a_t \uparrow \infty$ and $\sum_{t=1}^{n} \frac{x_t}{a_t} \longrightarrow c$ (a constant) as $n \longrightarrow \infty$.
$\Longrightarrow \frac{1}{a_n} \sum_{t=1}^{n} x_t \longrightarrow 0$ as $n \longrightarrow \infty$

1. Let $\{X_i, i \geq 1\}$ be a sequence of i.i.d. r.v.s with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2 \neq 0$. Then

$$\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

Note that no higher moment conditions are required.
2. Let $\{X_i, i \geq 1\}$ be a sequence of independent r.v.s with mean μ_i and variance σ_i^2, and let $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$. If

$$\frac{\max_{1\leq i\leq n}\left[E|X_i - \mu_i|^{2+\delta}\right]^{\frac{1}{2+\delta}}}{\bar{\sigma}_n} \leq B < \infty \quad (\delta > 0)$$

for all n,

$$\frac{\sum_{i=1}^{n}(X_i - \mu_i)}{\bar{\sigma}_n\sqrt{n}} \xrightarrow{d} N(0,1).$$

The central limit theorem also works for a sequence of serially dependent ramdom variables.
See Chapter 2 of Lehmann (1999; Elements of Large-Sample Theory) and Chapter 24 of Davidson (1994).

The sequence $\{X_i, i \geq 1\}$ is m-dependent if $(X_1, ..., X_i)$ and $(X_j, ..., X_n)$ are independent whenever $j - i > m$.
Suppose that $\{X_i, i \geq 1\}$ is $m-$dependent and stationary and let

$$\sigma^2 = Var(X_1) + 2\sum_{i=2}^{m+1} Cov(X_1, X_i).$$

Then,

$$\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1) \text{ as } n \rightarrow \infty.$$

Suppose that $\{X_i, i \geq 1\}$ is a sequence of i.i.d. r.v.s that follow the Cauchy distribution $C(0,1)$. The distribution of $\bar{X}$ is $C(0,1)$ for all n. The Cauchy distribution $C(a,b)$ is a symmetric, bell-shaped distribution on $(-\infty, \infty)$ with pdf

$$f(x \mid \theta) = \frac{b}{\pi} \frac{1}{b^2 + (x-a)^2}, \; -\infty < x < \infty, \; -\infty < \theta < \infty.$$

The mean of the Cauchy distribution does not exist, i.e., $E|X| = \infty$.

Let $X_1, ..., X_n$ be i.i.d. according to the Poisson distribution $P(\lambda)$ with $\lambda = \frac{1}{n}$. Then, $\sum_{i=1}^{n} X_i$ is distributed as $P(1)$, and hence ($\mu = \sigma^2 = 1/n$)

$$\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\right)}{\sqrt{\frac{1}{n}}\sqrt{n}} = \sum_{i=1}^{n} X_i - 1.$$

Thus, the CLT does not work. (Why? The underlying distribution depends on n.)

Example

Let $X_i \sim iidB(1, p)^a$. Then $EX_1 = p$ and $Var(X_1) = p(1-p)$. Thus,

$$\frac{\sum_{i=1}^{n}(X_i - p)}{\sqrt{p(1-p)}\sqrt{n}} \xrightarrow{d} N(0,1).$$

[a] This means X_i are independent having the Bernoulli distribution with success probability p.

- For vector sequences, we use the following result known as the Cramer–Wold device.
- If $\{X_n\}$ is a sequence of random vectors, $X_n \xrightarrow{d} X$ iff $\lambda' X_n \xrightarrow{d} \lambda' X$ for any vector $\lambda \in R^K - \{0\}$.

Example

Let $\underset{m\times 1}{X_i} \sim iid\,(0, \Sigma)$. Then,

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \xrightarrow{d} N(0, \Sigma).$$

Consistency of the least squares estimator

- Assume

1. (x_i, ε_i) is a sequence of independent observations.
2. $E(\varepsilon_i \mid x_1, ..., x_n) = 0$ for all i.
3. $\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \left(= \frac{1}{n}X'X\right) \xrightarrow{p} Q = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E(x_i x_i') \ (> 0)$.
4. For any $\lambda \in R^K - \{0\}$ and $\delta > 0$, $E\left|\lambda' x_i \varepsilon_i\right|^{2+\delta} \leq B < \infty$ for all i.

- The least squares estimator b may be written as

$$b = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i\right)$$

Consistency of the least squares estimator

- Consider for $\lambda \in R^K - \{0\}$

$$\frac{1}{n}\sum_{i=1}^{n} \lambda' x_i \varepsilon_i = \frac{1}{n}\sum_{i=1}^{n} w_i.$$

Then, w_i is an independent sequence with

$$E(w_i) = EE\left(\lambda' x_i \varepsilon_i | X\right) = 0.$$

In addition, Assumption 3 implies $E|w_i|^{1+\delta} \leq D < \infty$ for all i.

Consistency of the least squares estimator

Lyapounov's inequality For $0 < \alpha \leq \beta$, $\left(E|X|^{\alpha}\right)^{1/\alpha} \leq \left(E|X|^{\beta}\right)^{1/\beta}$.

Thus, by the WLLN for an independent sequence,

$$\frac{1}{n}\sum_{i=1}^{n} w_i \xrightarrow{p} 0.$$

Since this holds for any $\lambda \in R^K - \{0\}$,

$$\frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i \xrightarrow{p} 0$$

and we have

$$b \xrightarrow{p} \beta + Q^{-1} \cdot 0 = \beta.$$

Consistency of the least squares estimator

A condition for the consistency of b : Assume X is a nonstochastic matrix. Let

$$\lambda'(X'X)^{-1}X'\varepsilon = z.$$

Then $E(z) = 0$ and

$$Var(z) = \sigma^2\lambda'(X'X)^{-1}\lambda.$$

Due to Rayleigh-Ritz inequality,

$$\sigma^2\lambda'(X'X)^{-1}\lambda \leq ev_{\max}\left((X'X)^{-1}\right)\sigma^2\lambda'\lambda.$$

Thus, if $ev_{\max}\left((X'X)^{-1}\right) = \frac{1}{ev_{\min}(X'X)} \to 0$ or $ev_{\min}(X'X) \to \infty$, b is consistent for β.

Rayleigh-Ritz inequality For a real, symmetric matrix A,

$$ev_{\min}(A)x'x \leq x'Ax \leq ev_{\max}(A)x'x.$$

Asymptotic normality of the least squares estimator

- Write

$$b - \beta = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i \varepsilon_i$$

or

$$\sqrt{n}\,(b - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \varepsilon_i \right).$$

Since $\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} Q$ by assumption, we need to show that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \varepsilon_i$ is normally distributed in the limit.

Asymptotic normality of the least squares estimator

- Consider for $\lambda \in \mathbb{R}^K - \{0\}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda' x_i \varepsilon_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i .$$

  We need to check the conditions of the CLT for a sequence of independent r.v.'s.

1. $E\left(w_i\right) = 0$ as before.
2. $\left(E|w_i|^{2+\delta}\right)^{\frac{1}{2+\delta}} \leq B^{\frac{1}{2+\delta}}$ for all i and $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n Ew_i^2 \leq B$.

- Thus

$$\frac{\sum_{i=1}^n w_i}{\bar{\sigma}_n \sqrt{n}} \xrightarrow{d} N\left(0, 1\right).$$

Asymptotic normality of the least squares estimator

- Since

$$
\begin{aligned}
\bar{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^{n} E\left(w_i^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left(\lambda' x_i \varepsilon_i \varepsilon_i' x_i' \lambda\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} EE\left(\lambda' x_i \varepsilon_i \varepsilon_i' x_i' \lambda | X\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left(\lambda' x_i E\left(\varepsilon_i \varepsilon_i' | X\right) x_i' \lambda\right) \\
&= \frac{\sigma^2}{n}\lambda' \sum_{i=1}^{n} E(x_i x_i')\lambda \\
&\to \sigma^2 \lambda' Q \lambda,
\end{aligned}
$$

Asymptotic normality of the least squares estimator

- This result can be written as

$$\frac{\sum_{i=1}^{n} w_i}{\sqrt{n}} \xrightarrow{d} N\left(0, \sigma^2 \lambda' Q \lambda\right),$$

which implies

$$\frac{\sum_{i=1}^{n} x_i \varepsilon_i}{\sqrt{n}} \xrightarrow{d} N\left(0, \sigma^2 Q\right).$$

Asymptotic normality of the least squares estimator

- Using this and the given assumption, we have

$$\sqrt{n}\,(b-\beta) \xrightarrow{d} N\left(0, \sigma^2 Q^{-1}\right).$$

(Recall that $X_n Y_n \xrightarrow{d} cY$ if $X_n \xrightarrow{p} c$ (a constant) and $Y_n \xrightarrow{d} Y$)

- Some authors write this result as

$$b \simeq N\left(\beta, \frac{1}{n}\left(\sigma^2 Q^{-1}\right)\right).$$

That is, b is approximately normal with mean β and variance–covariance matrix $\frac{1}{n}\left(\sigma^2 Q^{-1}\right)$.

Consistency of estimators of the error variance

- Write

$$
\begin{aligned}
s^2 &= \frac{1}{n-K}\varepsilon' M \varepsilon \\
&= \frac{1}{n-K}\left[\varepsilon'\varepsilon - \varepsilon' X \left(X'X\right)^{-1} X'\varepsilon\right] \\
&= \frac{n}{n-K}\left[\frac{\varepsilon'\varepsilon}{n} - \frac{\varepsilon' X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\right].
\end{aligned}
$$

Consistency of estimators of the error variance

- Because

$$\frac{\varepsilon'\varepsilon}{n} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 \xrightarrow{p} \sigma^2$$

$$\frac{\varepsilon' X}{n} = \frac{1}{n}\sum_{i=1}^{n}x_i\varepsilon_i \xrightarrow{p} 0$$

$$\frac{X'X}{n} = \frac{1}{n}\sum_{i=1}^{n}x_i x_i' \xrightarrow{p} Q$$

  and

$$\frac{n}{n-K} \to 1,$$

$$s^2 \xrightarrow{p} \sigma^2.$$

Consistency of estimators of the error variance

- That is, σ^2 is consistently estimated by s^2. Alternatively, we may use

$$\hat{\sigma}^2 = \frac{1}{n} e' e.$$

  This is also consistent.

Asymptotic distribution of a function of b

- Let $f(b)$ be a vector of J continuous and continuously differentiable functions of b. We want to find the limiting distribution of $f(b)$. By the Taylor expansion

$$f(b) = f(\beta) + \frac{\partial f(\beta)}{\partial \beta'}(b - \beta) + \text{remainder}.$$

$\frac{\partial f(\beta)}{\partial \beta'}$ is a matrix of the form

$$\begin{bmatrix} \frac{\partial f_1(\beta)}{\partial \beta_1} & \cdots & \frac{\partial f_1(\beta)}{\partial \beta_K} \\ \vdots & & \\ \frac{\partial f_J(\beta)}{\partial \beta_1} & \cdots & \frac{\partial f_J(\beta)}{\partial \beta_K} \end{bmatrix} = \Gamma.$$

Asymptotic distribution of a function of b

- The remainder term becomes negligible if $b \xrightarrow{p} \beta$.
- Assume $\Gamma \neq 0$. Then,

$$\sqrt{n}\left(f\left(b\right)-f\left(\beta\right)\right) \xrightarrow{d} N\left(0, \Gamma\left(\sigma^2 Q^{-1}\right)\Gamma'\right).$$

- That is, $f\left(b\right)$ also has a normal distribution in the limit.

Asymptotic distribution of a function of b

Example

Suppose

$$\sqrt{n}\,(b-\beta) \xrightarrow{d} N\left(0, \sigma^2 Q^{-1}\right).$$

What is the distribution of $b_1^2 + \cdots + b_K^2$?
Here

$$f\,(b) = b_1^2 + \cdots b_K^2$$

and

$$\Gamma = \left[2\beta_1, \cdots, 2\beta_K\right].$$

Thus, if $\Gamma \neq 0$,

$$\sqrt{n}\left(b_1^2 + \cdots + b_k^2 - \left(\beta_1^2 + \cdots + \beta_k^2\right)\right) \xrightarrow{d} N\left(0, \sigma^2 \Gamma Q^{-1} \Gamma'\right).$$

More general assumption on the regressors

- We have assumed (X_i) is a sequence of independent observations. This assumption may occasionally be violated in practice. For example, consider the autoregressive model of order p

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t$$

  where $\varepsilon_t \sim iid\left(0, \sigma^2\right)$.
- Here, the regressors are correlated over time. Still, consistency and asymptotic normality of the OLS estimator follows if we make a few extra assumptions. These should be dealt with in a more advanced course.

Instrumental variables estimation

- We have assumed

$$E\left(\varepsilon_i | X\right) = 0$$

which implies $E\left(\varepsilon_i X_i\right) = 0$. There are many examples where this assumption is violated.

Instrumental variables estimation

Example

Simultaneous equations

Let C_t : consumption at time t

Y_t : income at time t

I_t : investment at time t

The Keynesian consumption function is

$$C_t = \alpha + \beta Y_t + \varepsilon_t.$$

But $Y_t = C_t + I_t$. Using these two equations, we have

$$Y_t = \alpha + \beta Y_t + \varepsilon_t + I_t \Rightarrow Y_t = \frac{1}{1-\beta}\left(\alpha + \varepsilon_t + I_t\right).$$

Thus Y_t and ε_t are correlated.

Instrumental variables estimation

Example

Autoregressive Moving Average model

Consider the ARMA$(1,1)$ model

$$y_t = \alpha y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\varepsilon_t \sim iid\left(0, \sigma^2\right)$$

$$|\alpha| < 1, |\theta| < 1.$$

Writing

$$y_t = \alpha y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\alpha y_{t-1} = \alpha^2 y_{t-2} + \alpha \varepsilon_{t-1} + \alpha \theta \varepsilon_{t-2}$$

$$\vdots$$

Example

(continued) and adding all of these equations, we obtain

$$y_t = \varepsilon_t + (\theta + \alpha)\,\varepsilon_{t-1} + \alpha\,(\theta + \alpha)\,\varepsilon_{t-2} + \cdots .$$

Thus y_{t-1} and ε_{t-1} are correlated.

Instrumental variables estimation

Example

Measurement error

Let the true regression model be

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

Suppose that we observe

$$x_i^* = x_i + w_i \ \left(w_i \sim iid\left(0, \sigma_w^2\right)\right)$$

instead of x_i due to measurement error. Then, the regression model we use will be

$$\begin{aligned} y_i &= \alpha + \beta\left(x_i^* - w_i\right) + \varepsilon_i \\ &= \alpha + \beta x_i^* + \varepsilon_i - \beta w_i. \end{aligned}$$

Instrumental variables estimation

Example

Dynamic panel data model

Let

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + u_{it}$$

$$u_{it} = \mu_i + v_{it} \ \text{(one–way error component model)}$$

where $\mu_i \sim iid\left(0, \sigma_\mu^2\right)$ and $v_{it} \sim iid\left(0, \sigma_v^2\right)$. We call μ_i unobserved individual effect variable. Since $y_{i,t-1}$ is a function of μ_i, it is correlated with u_{it}. Thus, OLS estimator is inconsistent.

Instrumental variables estimation

- Suppose that a sequence of $K \times 1$ vector, z_i, satisfies

$$\frac{1}{n}\sum_{i=1}^{n} z_i \varepsilon_i \xrightarrow{p} 0$$

and

$$\frac{1}{n}\sum_{i=1}^{n} z_i x_i' \xrightarrow{p} Q_{ZX}.$$

Then,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} z_i y_i &= \frac{1}{n}\sum_{i=1}^{n} z_i x_i'\beta + \frac{1}{n}\sum_{i=1}^{n} z_i \varepsilon_i \\ &\xrightarrow{p} Q_{ZX}\beta \end{aligned}$$

Instrumental variables estimation

- Thus, as an estimator of β, we consider

$$b_{IV} = \left(Z'X\right)^{-1} Z'y,$$

where $Z = \begin{bmatrix} z_1' \\ z_2' \\ \vdots \\ z_n' \end{bmatrix}$.

Instrumental variables estimation

- Assume

1. $E\left(\varepsilon|X\right) \neq 0.$
2. $\frac{1}{n}Z'X \xrightarrow{p} Q_{ZX}$ with $rank(Q_{ZX}) = K.$
3. $\frac{1}{n}Z'Z \rightarrow Q_{ZZ}\left(> 0\right).$
4. $E\left(\varepsilon|Z\right) = 0.$
5. $E\left(\varepsilon\varepsilon'|Z\right) = \sigma^2 I.$
6. $\left(z_i, \varepsilon_i\right)$ is a sequence of independent observations.
7. For any $\lambda \in R^K - \{0\}$and $\delta > 0$, $E\left|\lambda' z_i \varepsilon_i\right|^{2+\delta} \leq B < \infty$ for all i.

Instrumental variables estimation

- Write the IV estimator as

$$b_{IV} = \beta + \left(Z'X\right)^{-1} Z'\varepsilon.$$

Since $\frac{1}{n}Z'X \xrightarrow{p} Q_{ZX}$ and $\frac{1}{n}Z'\varepsilon \xrightarrow{p} 0$, $b_{IV} \xrightarrow{p} \beta$ as $n \longrightarrow \infty$.

- In addition,

$$\frac{1}{\sqrt{n}}Z'\varepsilon \xrightarrow{d} N\left(0, \sigma^2 Q_{ZZ}\right),$$

which implies

$$\sqrt{n}\left(b_{IV} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2 Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}\right).$$

Instrumental variables estimation

- As for the OLS estimation, a natural estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i' b_{IV}\right)^2 .$$

  We can show that

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2 \text{ as } n \longrightarrow \infty.$$

- In practice, the instruments and regressors may be only weakly correlated. If so, the variance of the IV estimator is large. Such instrumental variables are commonly called weak instruments.

Instrumental variables estimation

- So far, the number of instruments is equal to the number of regressors. What if number of instruments is greater than the number of regressors?
- Then we use

$$b_{IV} = \left[X'Z \left(Z'Z \right)^{-1} Z'X \right]^{-1} X'Z \left(Z'Z \right)^{-1} Z'y.$$

Instrumental variables estimation

- This is equivalent to

$$\left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'y$$

where

$$\hat{X} = Z\left(Z'Z\right)^{-1}Z'X.$$

$\hat{X}$ is the part of X explained by Z. Because Z is uncorrelated with the error terms, this part should also be uncorrelated with the error terms. The estimator is called the two-stage least squares estimator. Its asymptotic properties are:

1. $b_{IV} \xrightarrow{p} \beta$
2. $\sqrt{n}\left(b_{IV} - \beta\right) \xrightarrow{d} N\left(0, \sigma^2\left(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}\right)^{-1}\right)$

Overidentification test

- Suppose that there are K ($= \#$ of regressors) instruments. Can we check if these instruments are exogenous? (i.e., does $E\left(\varepsilon|Z\right) = 0$ hold?). The answer is: no.
- But if there are K exogenous instruments, we can test whether other additional instruments are exogenous. This can be done by the overidentification test.
- Let $e_{IVi} = y_i - \hat{b}_{IV}' x_i$. Regress e_{IVi} on z_i ($(K+m) \times 1$ vector) and test the significance of the coefficients by using the test statistic $J = (K+m)F$. This test statistic has $\chi^2(m)$ as its limiting distribution. If the null is not rejected, the additional instruments are valid.

Durbin–Wu–Hausman test

- See Durbin (1954, Review of the International Statistical Institute), Wu (1973, Econometrica), and Hausman (1978, Econometrica).
- The null hypothesis for the DWH test is

$$H_0 : E\left(X'\varepsilon\right) = 0$$

- Under H_0, $b_{IV} - b \xrightarrow{p} 0$. If H_0 is violated, $b_{IV} - b \xrightarrow{p} \delta\,(\neq 0)$. Thus, the DWH test is based on $d = b_{IV} - b$.

Durbin–Wu–Hausman test

- Since the asymptotic variance–covariance matrix of d is

$$\begin{aligned} & Asy.Var\left(b_{IV}\right) - Asy.Var\left(b\right) \\ = \ & \sigma^2\left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1} - \sigma^2\left(X'X\right)^{-1}, \end{aligned}$$

  the DWH test is defined by

$$DWH = d'\left[s^2\left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1} - s^2\left(X'X\right)^{-1}\right]^{-1} d.$$

- As $n \longrightarrow \infty$,

$$DWH \xrightarrow{d} \chi^2_K.$$

Durbin–Wu–Hausman test

- If only J elements of X are not correlated with the error term, the DWH test is based on the corresponding IV and OLS estimators and has the asymptotic distribution

$$DWH \xrightarrow{d} \chi^2_J.$$

Exercises

1. Prove relation (1) by using Chebyshev's inequality.
2. Consider the linear regression

$$y_t = \beta t + \varepsilon_t, \ \varepsilon_t \sim iid(0,1).$$

What is the asymptotic distribution of $\frac{1}{b}$?

3. Let b_1 and b_2 be the OLS estimates of of β_1 and β_2 of the standard linear regression model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, (i = 1, ..., n).$$

Derive the limiting distribution of $\sqrt{n}(\frac{b_1}{b_2} - \frac{\beta_1}{\beta_2})$.

4. Let the true regression model be

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

But we observe

$$x_i^* = x_i + w_i \ \ \left(w_i \sim iid\left(0, \sigma_w^2\right)\right)$$

instead of x_i due to measurement error. When y_i is regressed on x_i^*, what is the probability limit of the OLS estimator? Assume that $\{\varepsilon_i\}$ and $\{w_i\}$ are independent.

Exercises

5. Consider the simultaneous equation model

$$y_{1t} = \beta y_{2t} + \alpha z_{1t} + u_t. \tag{2}$$

The reduced form model for this model is written as

$$\begin{aligned} y_{1t} &= \pi_{11} z_{1t} + \pi_{12} z_{2t} + v_{1t}; \\ y_{2t} &= \pi_{21} z_{1t} + \pi_{22} z_{2t} + v_{2t}, \end{aligned}$$

where $(v_{1t}, v_{2t})' \sim (0, \Sigma)$. Derive the probability limit of the OLS estimator of β using the structural equation (2). Is it consistent?

6. In the linear regression

$$y_i = \beta' X_i + \varepsilon_i,$$

show that $Cov(b, b - b_{IV}) = 0$ when $E(X_i \varepsilon_i) = 0$ for all i.[1]

[1] This fact is utilized for the DWH test.

Exercises

7. Consider the linear regression model $y = X\beta + \varepsilon$ where X are all endogenous. There are K valid instrumental variables available. We want to test whether m additional instruments are exogenous. In order to test this, we regress $e_{IVi} = y_i - \hat{b}_{IV}' x_i$ on z_i ($(K + m) \times 1$ vector) and test the significance of the coefficients by using the test statistic $J = (K + m)F$. Show that this test statistic has $\chi^2(m)$ as its limiting distribution by following the steps below.

a. Let the regression coefficient obtained by regressing $e_{IVi} = y_i - \hat{b}_{IV}' x_i$ on z_i be denoted as $\hat{\gamma}$. Show that

$$\hat{\gamma} = (Z'Z)^{-1} Z' [I - X(X'P_z X)^{-1} X' P_z] \varepsilon.$$

Exercises

b. Show that the J statistic can be written as

$$\begin{aligned} J &= \hat{\gamma}'[\hat{\sigma}^2(Z'Z)^{-1}]^{-1}\hat{\gamma} \\ &= \varepsilon'[I - P_z X(X'P_z X)^{-1}X']P_z[I - X(X'P_z X)^{-1}X'P_z]\varepsilon/\hat{\sigma}^2 \\ &= w/\hat{\sigma}^2, \text{ say,} \end{aligned}$$

where $\hat{\sigma}^2 = (e_{IV} - Z\hat{\gamma})'(e_{IV} - Z\hat{\gamma})/n$.

c. Find the limiting distribution of w.

d. Find the probability limit of $\hat{\sigma}^2$.

e. Using parts c and d, find the limiting distribution of the J statistic.

8.Suppose that z_i (a scalar variable) is a valid instrument for x_i in the model

$$y_i = \beta' x_i + \varepsilon_i.$$

One considers the following OLS regression

$$\hat{y}_i = \hat{\beta}' x_i + \hat{\gamma} w_i,$$

where w_i is the OLS regression residual obtained by regressing x_i on z_i. How is $\hat{\beta}$ related to the usual IV estimator that uses z_i as an instrument?

9. In a regression model

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

the regressor and errors are correlated. But a valid instrument z_i is available. Consider switching the independent and dependent variables and running an IV regression using z_i as an instrument. Is the IV estimator consistent? What is its limiting distribution?

10. In a linear regression model, the error terms are heteroskedastic and correlated with the regressors. How can you test linear null hypotheses on the slope coefficients properly? Assume that valid instruments are available.

11. For the linear regression model

$$y_t = \beta x_t + \varepsilon_t, \ \ \varepsilon_t \sim iid\left(0, \sigma^2\right), \ \ t = 1, \cdots, n,$$

an estimator

$$\bar{b} = \frac{\sum_{t=1}^{n} y_t}{\sum_{t=1}^{n} x_t}$$

is considered. Assume $\{X_t\}$ is a sequence of constants.

a. What assumptions are required for the consistency of $\bar{b}$.

b. Derive the asymptotic distribution of $\bar{b}$.

Exercises

12.Consider the linear regression model

$$y_i = \frac{\beta}{x_i} + \varepsilon_i,\ \ \varepsilon_i \sim iid\left(0, \sigma^2\right),\ \ i = 1, \cdots, n$$

when $\{X_i\}$ is a sequence of constants. What assumptions are required for the consistency of the LSE of β? Are these assumptions reasonable for empirical analysis?

13. Consider the linear regression model

$$y_t = \frac{\beta}{t} + \varepsilon_t,\ \ \varepsilon_t \sim iid\left(0, \sigma^2\right),\ \ t = 1, \cdots, n.$$

Is the OLS estimator of β consistent?

Exercises

14. Consider the standard linear regression model

$$y_i = \beta x_i + \varepsilon_i, \ (i = 1, ..., n),$$

where $\{x_i, \varepsilon_i\}$ is a sequence of independent observations, $E(\varepsilon_i \mid x_i) = 0$, $E(\varepsilon_i \mid x_i) = \sigma^2$ and $E|\varepsilon_i|^{2+\delta} \leq B < \infty$ ($\delta > 0$) for all i. One considers using the IV estimator

$$\tilde{\beta} = \frac{\sum_{i=1}^{n} sgn(x_i) y_i}{\sum_{i=1}^{n} sgn(x_i) x_i}$$

to estimate the slope coefficient β.

a. Show that $\tilde{\beta}$ is consistent if $\frac{1}{n}\sum_{i=1}^{n}|x_i| \xrightarrow{p} c$ (a postive constant).

b. Derive the limiting distribution of $\tilde{\beta}$.

c. Derive the limiting distribution of $\frac{1}{\tilde{\beta}}$.

d. Construct the t-ratio using $\tilde{\beta}$ and derive its limiting distribution.

Exercises

15. Consider the standard linear regression model

$$y_i = \beta x_i + \varepsilon_i, \ \ (i = 1, ..., n) \, .$$

It is known that

$$\sqrt{n} \left(\hat{\beta}_{ols} - \beta \right) \xrightarrow{d} N(0, \sigma^2 Q^{-1}),$$

where $Q = p \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i^2$.

a. What is the limiting distribution of $\sqrt{n} \left(\left| \hat{\beta}_{ols} \right| - |\beta| \right)$ when $\beta \neq 0$?

b. What is the limiting distribution of $\sqrt{n} \left(\left| \hat{\beta}_{ols} \right| - |\beta| \right)$ when $\beta = 0$?

Exercises

16. Consider the linear regression model

$$y_t = \beta' x_t + \varepsilon_t, \ (t = 1, ..., T),$$

where $\varepsilon_1, \ \varepsilon_2, ...$ are i.i.d. with $E(\varepsilon_t) = 0, \ Var(\varepsilon_t) = \sigma^2$ and $Var(\varepsilon_t^2) = \tau$. Suppose that $\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \xrightarrow{p} Q \ (> 0)$ and that $\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t \xrightarrow{d} N(0, \sigma^2 Q)$. We consider estimating σ^2 by $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2$, where $e_t = y_t - b' x_t$. Derive the limiting distribution of $\sqrt{T}(\hat{\sigma}^2 - \sigma^2)$.