Large Sample Results for the Linear Model

Walter Sosa-Escudero

Econ 507. Econometric Analysis. Spring 2009

February 19, 2009

Why Large Sample Theory?

- $\hat{\beta} = (X'X)^{-1}X'Y$, then $E(\hat{\beta}) = \beta$. Easy, mostly due to linearity.
- What about $\hat{\beta}^* = g(X, Y)$. In many cases, impossible to derive finite sample properties without being too specific about $g(.)$.

- Consider $H_0 : \beta = 0$ vs. $H_A : \beta \neq 0$. We relied on checking $\hat{\beta} \cong 0$, based on the distribution of $\hat{\beta}$. We had to assume u normal and, through linearity, we got the distribution of $\hat{\beta}$.

- It is equivalent to think about $H_0 : \delta\beta = 0$ vs. $H_A : \delta\beta \neq 0$ for $\delta \neq 0$ and check whether $\delta\hat{\beta} \cong 0$. Why? If we choose δ cleverly, the distribution of $\delta\hat{\beta}$ is much easier to get than that of $\hat{\beta}$. For example, if we set $\delta = \sqrt{n}$, when $n \to \infty$, $\delta\hat{\beta} \simeq N$.

Example: $X \sim (\mu, \sigma^2)$, $H_0 : \mu = \mu_0$, $S^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$

X **normal:**

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} \overset{a}{\simeq} t_{n-1}$$

Reject if $|z| > c_\alpha$, with $Pr(|Z| > c_\alpha) = \alpha$.

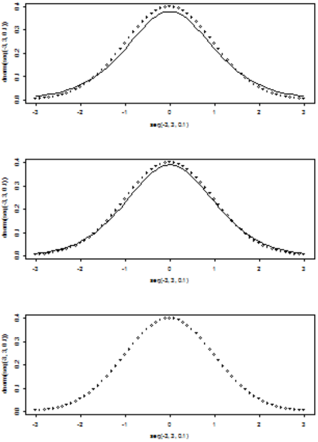
X **not normal:**

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{S^2/n}} \simeq N(0,1)$$

Reject if $|z| > c_\alpha$, with $Pr(|Z| > c_\alpha) = \alpha$.

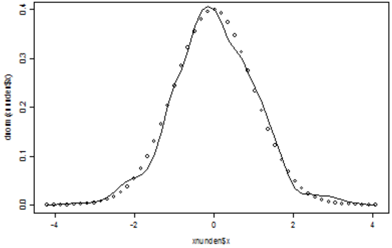
- If X is normal, we use the t distribution. The result holds for any sample size.
- If X is not normal, we use the *normal* distribution. The result holds asymptotically.

- What do we mean by 'valid asymptotically'?
- How relevant is the previous result in practice, when n is necesarily finite



X is normal

- We should use the t distribution to compute c_α.
- What if we use the normal?
- The approximation works fine. For $n = 30$ differences are negligible.



X is not-normal

- Here we do not have information to compute c_α.
- What if we use the normal?
- Uniform case: the approximation works well!

- Large sample approximations work well in many cases, even far from infinity.
- In most cases it is much easier to derive the asymptotic approximations instead of the finite sample result.

Basic Concepts

Convergence in probability:

A sequence $\{z_n\}$ of random scalars *converges in probability* to a non-random constant a if for every $\epsilon > 0$:

$$\lim_{n \to \infty} Pr\left[|z_n - a| > \epsilon\right] = 0$$

- We use the notation $z_n \xrightarrow{p} a$, or *plim* $z_n = a$
- It extendes naturally to sequences of random vectors or matrices, requiring element-by-element convergence.

Almost sure convergence:

A sequence $\{z_n\}$ of random scalars *converges almost surely* to a constant a if

$$Pr\left(\lim_{n\to\infty} z_n = \alpha\right) = 1$$

- We use the notation $z_n \overset{a.s.}{\to} a$, or *plim* $z_n = a$
- Almost sure convergence implies convergence in probability.

Convergence in Distribution A sequence $\{z_n\}$ of random scalars with distribution cumulative distribution F_n *converges in distribution* to a random scalar z with distribution F

$$\lim_{n\to\infty} F_n = F$$

at every continuity point of F.

- F is the *limiting distribution* of $\{z_n\}$.
- Notation: $z_n \xrightarrow{d} z$.

Law of Large Numbers

Let z_n be a secuence of random scalars, and construct a new sequence

$$\bar{z}_n = \frac{\sum_{i=1}^{n} z_i}{n}$$

Kolmogorov's Strong LLN: $\{z_n\}$ and i.i.d. secuence of random scalars with $E(z_i) = \mu$ (exists and is finite). Then $\bar{z}_n \xrightarrow{a.s.} \mu$.

Central Limit Theorem

Lindeberg-Levy CLT: $\{z_n\}$ and i.i.d. secuence of random scalars with $E(z_i) = \mu$ and $V(z_i) = \sigma^2$ (both exist and are finite). Then

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

- Careful, the theorem does not refer to $\bar{z}_n$ but to a transformation involving n.

Useful Results

- Continuity: $\{z_n\}$ a sequence of random vectors, z a random vector, a a vector of constants. Let $g(.)$ be a vector valued continuous function that does not depend on n
  - $z_n \xrightarrow{p} a \;\Rightarrow\; g(z_n) \xrightarrow{p} g(a)$, and $z_n \xrightarrow{a.s.} a \;\Rightarrow\; g(z_n) \xrightarrow{a.s.} g(a)$
  - $z_n \xrightarrow{d} z \;\Rightarrow\; g(z_n) \xrightarrow{d} g(z)$
  - Product rule: $z_n \xrightarrow{p} 0$ and $x_n \xrightarrow{d} x$, then $z_n x_n \xrightarrow{p} 0$
- Slutzky's Theorem:
  - $x_n \xrightarrow{d} x$ and $y_n \xrightarrow{p} \alpha$, then $x_n + y_n \xrightarrow{d} x + \alpha$.
  - $x_n \xrightarrow{d} x$, $A_n \xrightarrow{p} A$, then $A_n x_n \xrightarrow{d} Ax$, provided A_n and x_n are conformable.

- Cramer Wold Device (restricted): x_n a sequence of $K \times 1$ random vectors. If for any vector $\lambda \in \Re^K$, $\lambda' x_n \xrightarrow{d} N$ then x_n converges to a multivariate normal random variable.
- Asymptotic equivalence: If $(x_n - y_n) \xrightarrow{p} 0$ and $y_n \xrightarrow{d} Z$, then $x_n \xrightarrow{d} Z$. We will say that x_n and y_n are *asymptotically equivalent.*
- $\{x_n, u_n\}$, i.i.d., then $\{x_i x_i'\}$ and $\{x_i u_i\}$ are iid.

Estimators as sequences of random variables

An estimator $\hat{\theta}_n$ is any function of the sample. The sequence $\{\hat{\theta}_n\}_n$ refers to the sequence formed by increasing the sample size progressively.

- Consistency: $\hat{\theta}_n$ is consistent for θ_0 if $\hat{\theta}_n \rightarrow \theta_0$ ('p'is *weak* and 'a.s.' *strong* consistency.
- Asymptotic normality: $\hat{\theta}_n$ is *asymptotically normal* if $\sqrt{n}(\hat{\theta}_n - \theta_0) \stackrel{d}{\rightarrow} N(0, \Sigma)$.
- Σ is the asymptotic variance. Careful with this!
- Such estimators are usually called $\sqrt{n}$-consistent.

Large Sample Properties of OLS estimators

Aside: some new notation

Let us write our linear model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + u_i, \qquad i = 1, \ldots, n$$

as: $y_i = x_i' \beta + u_i$

where x_i is an $K \times 1$ vector $(x_{1i}, x_{2i}, \ldots, x_{Ki})'$.

Note that $x_i x_i'$ is a $K \times K$ *matrix*. It is easy to check

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

$\hat{\beta}_n$ refers to the sequence of OLS estimators formed by increasing the sample size.

The Model

Assumptions for asymptotic analysis

1. Linearity: $y_i = x_i'\beta_0 + u_i \qquad i = 1, \ldots, n.$
2. Random sample: $\{x_i, u_i\}$ is a jointly i.i.d. process.
3. Predeterminedness: $E(x_{ik}u_i) = 0$ for all i and $k = 1, \ldots, K$.
4. Rank condition: $\Sigma_x \equiv E(x_i x_i')$ finite positive definite.
5. $V(x_i u_i) = S$ finite positive definite.

The results

- Consistency: $\hat{\beta}_n \xrightarrow{p} \beta_0$.
- Asymptotic normality: $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{p} N(0, \Sigma_x^{-1} S \Sigma_x^{-1})$

Plan: first prove results, then we discuss the assumptions.

Consistency

$$\begin{aligned} \hat{\beta}_n &= \beta_0 + (X'X)^{-1}X'u \\ &= \beta_0 + \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'u}{n}\right) \end{aligned}$$

We will show:

- $n^{-1}X'u \xrightarrow{p} 0$
- $\left(\frac{X'X}{n}\right)^{-1}$ does not explode

This argument will appear several times in this course!

The $h-$th element of $\left(\frac{X'u}{n}\right)$ is

$$\frac{1}{n}\sum_{i=1}^{n} x_{hi}u_i = \frac{1}{n}\sum_{i=1}^{n} z_i, \qquad z_i \equiv x_{hi}u_i$$

with $E(z_i) = 0$, so by Kolmogorov's LLN

$$\frac{1}{n}\sum_{i=1}^{n} z_i = \frac{\sum_{i=1}^{n} x_{hi}u_i}{n} \xrightarrow{p} E(x_{hi}u_i) = 0$$

hence

$$\left(\frac{X'u}{n}\right) \xrightarrow{p} 0$$

The (h,j) element of $\frac{X'X}{n}$ is $\frac{\sum_{i=1}^{n} x_{hi}x_{ji}}{n}$. Since $E(x_{hi}x_{ji}) = \Sigma_{x,hj}$, which exists and is finte, by the LLN (element-by-element)

$$\frac{X'X}{n} \xrightarrow{p} \Sigma_x$$

Since Σ_x is invertible and by continuity:

$$\left(\frac{X'X}{n}\right)^{-1} \xrightarrow{p} \Sigma_x^{-1}$$

Summarizing:

$$\hat{\beta}_n = \beta_0 \; + \; \underbrace{\left(\frac{X'X}{n}\right)^{-1}}_{\xrightarrow{p} \Sigma_x^{-1}} \; \underbrace{\left(\frac{X'u}{n}\right)}_{\xrightarrow{p} 0}$$

Hence, by the product rule: $\hat{\beta}_n \xrightarrow{p} \beta_0$

Asymptotic normality

The starting point is now:

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'u}{\sqrt{n}}\right)$$

We have already shown:

$$\left(\frac{X'X}{n}\right)^{-1} \xrightarrow{p} \Sigma_x^{-1}$$

Now the goal is to find the asymptotic distribution of $\left(\frac{X'u}{\sqrt{n}}\right)$ and use our continuity results.

$$\left(\frac{X'u}{\sqrt{n}}\right) = \sqrt{n}\frac{X'u}{n}$$

is a vector of K VA's. By the Cramer-Wold Device, we will find the distribution of:

$$\sqrt{n}\; c'\left(\frac{X'u}{n}\right)$$

for every vector $c \in \Re^K$. Note:

$$\sqrt{n}\; c'\left(\frac{X'u}{n}\right) = \sqrt{n}\; c'\frac{\sum x_i u_i}{n} = \sqrt{n}\frac{\sum c' x_i u_i}{n} \equiv \sqrt{n}\;\frac{\sum z_i}{n}$$

with $z_i \equiv c' x_i u_i$, a scalar random variable

It is easy to check (do it as exersise) that the assumptions imply:

- $E(z_i) = 0$
- $V(z_i) = c'Sc < \infty$.

Then, by the Lindeberg-Levy CLT applied to $\sqrt{n}\bar{z}$:

$$\sqrt{n}\,(\bar{z} - 0) = c'\left(\frac{X'u}{\sqrt{n}}\right) \xrightarrow{d} N(0, c'Sc)$$

hence by the Cramer-Wald device:

$$\left(\frac{X'u}{\sqrt{n}}\right) \xrightarrow{d} N(0, S)$$

Then:

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = \underbrace{\left(\frac{X'X}{n}\right)^{-1}}_{\xrightarrow{p} \Sigma_x^{-1}} \quad \underbrace{\left(\frac{X'u}{\sqrt{n}}\right)}_{\xrightarrow{d} N(0,S)}$$

By Slutzky's theorem and by the linearity property of multivariate normal variables:

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{p} N(0, \Sigma_x^{-1} S \Sigma_x^{-1})$$

A useful simplification

If we further assume $E(u_i^2|x_i) = \sigma_0^2$, then using LIE:

$$S = V(x_i u_i) = E(x_i u_i u_i x_i') = \sigma_0^2 E(x_i x_i') = \sigma_0^2 \, \Sigma_x$$

So the asymptotic variance of $\hat{\beta}_n$ reduces to

$$\Sigma_x^{-1} S \Sigma_x^{-1} = \sigma_0^2 \Sigma_x^{-1} \Sigma_x \Sigma_x^{-1} = \sigma_0^2 \Sigma_x^{-1}$$

Comments on the assumptions

1. *Predeterminedness vs. Mean independence:* $E(x_{ik}u_i) = 0$ is an orthogonality assumption. If there is a constant in the model, $E(u_i) = 0$ and this implies $Cov(x_i, u_i) = 0$. $E(u_i|x_{ik}) = 0$ is a stronger assumption, since it implies that u_i is uncorrelated to *any* (measurable) function of x_i.
2. *Predeterminedness vs. Strict Exogeneity:* our old assumption was $E(u|X) = 0$, which implies $E(u_iX_i) = 0$. We are requiring a bit less than we did to show unbiasedness (careful).

Example: $Y_t = \beta_1 + \beta_2 Y_{t-1} + u_t,\ \ t = 1, 2, ...\ E(u_t Y_{t-1}) = 0$ (consistency), but $E(u|Y_{-1}) \neq 0$ (why?). The OLS estimator will be consistent but biased!

3. *Rank condition as no multicollinearity in the limit:* this guarantees the invertibility of $n^{-1}\ X'X$ in the limit.
4. *Rank condition requires variability in* X The rank condition implies

$$\frac{1}{n}X'X \xrightarrow{p} E(x_i x_i'), \text{finite pd.}$$

Consider the following model:

$$y_i = \beta_0 + \beta_1 \frac{1}{i} + u_i, \ \ i = 1, \ldots, n$$

Here the explanatory variable provides less information as $n \rightarrow \infty$.

5. *We are not ruling out conditional heteroskedasticity!:* the iid assumption implies $E(u_i^2)$ is constant (unconditional homoskedasticity), but the error term *can* be conditionally heteroskedastic.
6. *The intercept.* Note that

$$Cov(X_k, u_i) = E(X_{ki}u_i) - E(X_{ki})E(u_i)$$

   If there is a constant in the model, $E(X_{1i}u_i) = E(u_i) = 0$. Then, this joint with predeterminedness implies no contemporaneous correlation between explanatory variables and the error term.
7. Do not confuse consistency with unbiasedness.

Extensions

- The iid case: No fixed regressors, heteroskedasticity, dependences.
- Heterogeneity: moment restrictions (Markov, Lyapunov CLT).
- Dependences: martingale difference, ergodic theorem, etc.

Consistency of S^2

Result: Under Assumptions 1-4, $s^2 \xrightarrow{p} E(u_i^2)$, provided $E(u_i^2)$ exists and is finite.

Proof:

$$s^2 = \frac{e'e}{n-K} = \frac{n}{n-K}\frac{e'e}{n} = \frac{n}{n-K}\frac{u'Mu}{n}$$

Now

$$u'Mu = u'(I - X(X'X)^{-1}X')u$$

I'll let you finish the proof in the homework.