OLS - asymptotics

Nathaniel Beck

Department of Politics, NYU, New York, NY 10012, nathaniel.beck@nyu.edu

QII - Week 1

Plim and Consistency

- Convergence in probability
- Consider a sequence of random variables index by N
- plim $b_N = b$ means

$$\lim_{N \to \infty} \Pr[|b_N - b| < \epsilon] = 1 \quad (1)$$

- Now consider some scalar parameter, β, and estimator of it, $\hat{\beta}_N$ which is indexed by the sample size with which it is estimated
- This estimator is CONSISTENT if plim $\hat{\beta}_N = \beta$
- If an estimator is consistent, for large enough N the probability of the estimator being at all different from the parameter it is estimating goes to zero
- Thus for large enough N, the distribution of $\boldsymbol{\beta}_N$ collapses to a "spike" at the true β
- Consistency implies "asymptotically unbiased" but NOT vice versa!
- Consistency is hobgoblin of small minded econometricians
- Lots of bad estimators are consistent (eg for mean, average of every thousandth observation)

- Other forms: mean square convergence (expectation of squared distance from true value goes to zero) - stronger (that is it implies convergence in probability) but often easier to prove
- Convergence almost surely - more complicated, not all that important in practice
- For vector of parameters, same ideas, just need a more complicated measure of how far estimate is from parameter
- Use $(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}')$ or its square root

Law of large numbers

- 'Weak" if plim, "strong" if almost surely, we use weak only
- Let x_i be a set of random variables ($i = 1, \ldots, N$)
- With common mean μ (can generalize)
- $\text{plim}\,\overline{x_N} = \lim E(\overline{x_N}) = \mu$
- Easily true under stringent conditions
- x_i is INDEPENDENT IDENTICALLY DISTRIBUTED (iid) iid with finite expectation
- True under less stringent conditions
- The x_i can have different distributions if still independent, so long as some other fairly non-stringent conditions hold

OLS is consistent

- From before

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon. \qquad (2)$$

- need to show that the plim of second term is zero
- Note that as N gets large, $\mathbf{X}'\mathbf{X}$ gets large (but we are inverting) as does $\mathbf{X}'\varepsilon$ since both are being summed over N

- We thus use the trick of multiplying the first term by N and the second term by N^{-1}
- and then noting that the plim of a product is the product of the plim's (Slutsky's theorem)
- We need to assume that $(N^{-1}\mathbf{X}'\mathbf{X})$ is both finite and non-singular (note that $(N^{-1})^{-1} = N$)
- (the independent variables are sufficiently bounded and no-colinearity)
- Thus the first plim is finite
- The second plim is zero, by the law of large numbers, since $E(\mathbf{x}_i\varepsilon) = 0$ (by assumption)
- Thus OLS is consistent
- CAT - weaker LLN shows consistency under heteroskedasticity

Chebyshev Inequality

- VERY LOOSE BOUND BECAUSE HOLDS FOR ANY DIST
- For ANY RV Z with mean μ and variance σ^2

$$P(\|Z-\mu\| > k) \leq \frac{\sigma^2}{k^2} \tag{3}$$

- When doing simulations, choose a distribution with long tails
- A nice easy one is a Bernoulli with $p = .5$

Central Limit Theorem

- A sequence of random variables CONVERGES IN DISTRIBUTION to another random variable
- if at every point of continuity, c, $F_N(c)$ converges to $F(c)$
- F is called the asymptotic distribution
- CENTRAL LIMIT THEOREM - convergence in distribution of mean to a normal
- Let $\frac{\bar{X}_N - E(\bar{X}_N)}{\sqrt{\text{Var}(\bar{X}_N)}} = Z_N$
- CLT says that under certain conditions on the x_i guarantees that Z_N converge in distribution to a standard normal
- If it does so we say that Z_N is ASYMPTOTICALLY (STANDARD) NORMAL - $Z_N \sim AN(0,1)$
- Simple condition: if the x_i are iid, then CLT holds
- If independent but not identically distributed, reasonable conditions still assure CLT

Asymptotic normality of OLS

- Note that consistency means that OLS converges to a spike
- Want to know what happens on the way, so spread out distribution
- Multiply by $\sqrt{N}$
- Want limit distribution of

$$\sqrt{N}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = (N^{-1}\mathbf{X}'\mathbf{X})^{-1}\frac{1}{\sqrt{N}}\mathbf{X}'\varepsilon \quad (4)$$

Asymptotically unbiased

- By assumption the first term just converges to a some PD matrix
- The second term is just the sum of a bunch of random variables
- So by CLT, is normal
- The expectation of each term in the sum is zero (by assumption that the $\mathbf{X}_i$ terms are uncorrelated with the errors
- So the asymptotic normal distribution has mean zero
- or the mean of the asymptotic distribution of $\hat{\boldsymbol{\beta}}$is $\boldsymbol{\beta}$
- Not a shock since OLS is unbiased even in finite samples
- Many estimators that are biased are asymptotically unbiased
- Consider the biased estimator of $\sigma^2 = \frac{\sum_{i=1}^{N} e_i^2}{N}$

Asymptotic Variance

- To get the variance, note that by ASSUMPTION
- $N^{-1}\mathbf{X}'\mathbf{X}$ converges to a Positive Definite (non-singular) matrix
- (Assumes the **X** matrix is bounded, in that as we add more observations but divide by N, we remain finite)
- Trivial assumption for cross-sections
- Thus the variance is driven by the variance of the middle term
- By usual variance argument dividing a random variable by $\sqrt{N}$ divided the variance by N and premultiplying by **X** means we pre and post-multiply variance by **X**' and **X**, respectively
- Since asymptotically the data variances and covariances approach their expectation (this is law of large numbers)
- Thus the plim of $\frac{1}{\sqrt{N}}\mathbf{X}'\varepsilon$ is $E(\mathbf{X}'\mathbf{\Omega}\mathbf{X})$

- Under assumptions of homoskedasticity the middle term is just $\sigma^2 I$, else it is a diagonal matrix
- Under homoskedasticity the variance of the normal is just the usual OLS variance
- With heteroskedasticity is the White formula
- Thus OLS $\hat{\beta}$ is asymptotically normal with mean β and OLS variance

Finite Sample Properties

- While we often worship at the alter of consistency, and at a minimum demand that all estimators that are worthy of consideration are unbiased
- Consistency is the hobgoblin of many bad estimators (do OLS on 1% of the observations)
- Though inconsistency is usually bad
- And unbiasedness simply makes proofs easier and only useful if you are going to average a bunch of estimates (even though by definition you only get one)
- What is useful is Mean Squared Error
- $E((\hat{b} - b)^2)$ (for scalar, or take matrix analogue in various ways)
- If estimator is unbiased, this is just the variance of the estimator
- But, in general, MSE is just the sum of squared bias and variance
- So a lower variance biased estimator may be better (lower MSE) than some biased estimators

- In general no BEST (lowest MSE) estimator exists
- Thus note the GM theorem only guarantees B in the class of LU estimators
- Consistency just tells that we eventually get to a spike
- Question is when
- For finite samples, usually need simulation