Asymptotics

Charlie Gibbons
ARE 212

Spring 2011

Overview of applications of asymptotics

There are two classes of questions in econometrics:

- Estimation
  - Finding unbiased estimators
  - Law of large numbers; convergence in probability or almost surely
  - Consistent estimators may be biased, but the bias goes to 0 as n gets large
- Inference
  - Hypothesis testing, confidence intervals, "significance"
  - Central limit theorem; convergence in distribution
  - Gives distribution of estimator without placing (many) distributional assumptions on y or, equivalently, ϵ

Unbiasedness

We start by considering the LCE model.

We do *not* need asymptotics for unbiasedness of our estimator:

$$\begin{aligned}
\mathbb{E}[\hat{\beta}] &= \mathbb{E}\left[\left(X'X\right)^{-1}X'y\right] \\
&= \mathbb{E}\left[\left(X'X\right)^{-1}X'\left(X\beta+\epsilon\right)\right] \\
&= \beta + \mathbb{E}\left[\left(X'X\right)^{-1}X'\mathbb{E}[\epsilon \mid X]\right] \\
&= \beta.
\end{aligned}$$

We also get unbiasedness if we have the LPM with non-stochastic X and $\mathbb{E}[\epsilon] = 0$ (*i.e.*, we include an intercept term).

In the LPM with stochastic regressors, we can't show unbiasedness—we need to show consistency.

Convergence almost surely

Let ω be an infinite sequence of random variables/vectors. Let $b_n(\omega)$ be some function applied to the first n of them. Then, the sequence $\{b_n(\omega)\}$ *converges almost surely* or *almost everywhere* to b if

$$\Pr\left[\omega\colon \lim_{n\to\infty} b_n(\omega) = b\right] = 1;$$

i.e., the probability of finding a sequence of random variables such that the limit of the entire sequence $\{b_n\}$ does not converge to b is 0.

This is called *strong consistency*.

Convergence in probability

Using the same notation as above, the sequence $\{b_n(\omega)\}$ *converges in probability* to b if

$$\lim_{n\to\infty} \Pr\left[\omega\colon |b_n(\omega) - b| < \epsilon\right] = 1 \;\forall\, \epsilon > 0;$$

alternatively written,

$$\operatorname*{plim}_{n\to\infty} b_n(\omega) = b;$$

i.e., the probability of finding a sequence of random variables such that b_n (the function applied to the first n of the sequence, as opposed to the entire sequence) is "far" from b goes to 0.

This is *(weak) consistency*.

Consistency results

$$X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{\text{p}} X$$

If $X_n \xrightarrow{\text{a.s.}} X$ and $Y_n \xrightarrow{\text{a.s.}} Y$, then $X_n + Y_n \xrightarrow{\text{a.s.}} X + Y$ and $X_n Y_n \xrightarrow{\text{a.s.}} XY$.

More generally, if a function $g(\cdot)$ is continuous at X, then $g(X_n) \xrightarrow{\text{a.s.}} g(X)$ (the *continuous mapping theorem*).

The CMT should be surprising; recall that $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$.

Analogous results hold for convergence in probability.

Application of the CMT

Let $X_n \xrightarrow{\text{p}} X$ and $Y_n \xrightarrow{\text{p}} Y$. Then, by the continuous mapping theorem,

$$\frac{X_n}{Y_n} \xrightarrow{\text{p}} \frac{X}{Y},$$

though, in general,

$$\mathbb{E}\left[\frac{X_n}{Y_n}\right] \neq \frac{X}{Y};$$

The ratio of two sequences is consistent, but typically biased.

Komolgorov (strong) law of large numbers

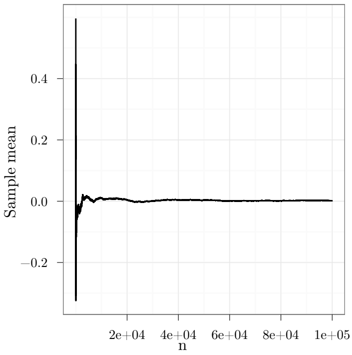
Let $\{Z\}$ be a sequence of i.i.d. random variables. Then $\bar{Z}_n \xrightarrow{\text{a.s.}} \mu$ if and only if $\mathbb{E}[Z] = \mu$ and $\mathbb{E}[|Z|] < \infty$, where

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^{n} Z_i.$$

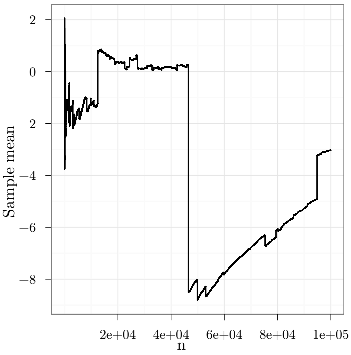
To provide some intuition, a sufficient condition for $\mathbb{E}[|Z|] < \infty$ is that $\mathrm{Var}(Z) < \infty$.

Proof of the weak LLN (*i.e.*, based upon convergence in probability, rather than almost-surely) is trivial using Chebyshev's inequality.

Finite variance assumption



(a) Normal distribution



(b) Cauchy distribution

Figure: Sequences of the sample mean with and without finite variance

Facts about i.i.d. variables

Let X and Y be identically distributed and $g(\cdot)$ be a continuous function. Then $g(X)$ and $g(Y)$ are identically distributed.

Instead, let X and Y be independent. Then $g(X)$ and $g(Y)$ are independent.

Linear projection model

Now, turn to the LPM with stochastic X. We have

$$\begin{aligned}\hat{\beta} &= \left(X'X\right)^{-1} X'y \\ &= \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i y_i\end{aligned}$$

and, using our modeling assumption,

$$\begin{aligned}\hat{\beta} &= \left(X'X\right)^{-1} X'y \\ &= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i.\end{aligned}$$

Consistency in the LPM

The law of large numbers implies that

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{\text{a.s.}} \mathbb{E}[xx'],$$

meaning that our model is asymptotically identified if the $\mathbb{E}[x'x]$ matrix is invertible, and that

$$\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{\text{a.s.}} \mathbb{E}[x\epsilon] = 0,$$

which gives consistency. We need a technical assumption as well, given on the coming summary slide.

Two derivations of consistency

We see consistency in two ways:

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i y_i \xrightarrow{\text{p}} \mathbb{E}\left[x x' \right]^{-1} \mathbb{E}\left[x y \right] = \beta$$

or

$$\hat{\beta} - \beta = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{\text{p}} \mathbb{E}\left[x x' \right]^{-1} \mathbb{E}\left[x \epsilon \right] = 0.$$

Assumptions for consistency in the LPM

We have a consistent model if we assume that

- $y = X\beta + \epsilon$,
- $\{(x_i, \epsilon_i)\}$ is an i.i.d. sequence,
- $\mathbb{E}[x_i \epsilon_i] = 0$,
- $\mathbb{E}[x_i x_i']$ is invertible, and
- the fourth moments of all the random variables exist (*i.e.*, are less than ∞).

We show that the stochastic LPM is consistent, not because it is biased, but because we can't deal with the term

$$\mathbb{E}\left[\left(X'X\right)^{-1} X'\epsilon\right]$$

in the proof of unbiasedness without additional assumptions.

An estimator for sample variance

We just saw an example showing that we may be able to demonstrate consistency when we can't say anything definitive about bias. But we also may show consistency *in spite of bias*.

Suppose that we use s^2 as an estimator for the sample variance of a random variable where

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

A biased estimator

$$\begin{aligned}
\mathbb{E}[s^2] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left((x_i-\mu)-(\bar{x}-\mu)\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(x_i-\mu)^2\right] - 2\frac{1}{n}\mathbb{E}\left[(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu)\right] \\
&\quad + \mathbb{E}\left[(\bar{x}-\mu)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(x_i-\mu)^2\right] - \mathbb{E}\left[(\bar{x}-\mu)^2\right] \\
&= \mathrm{Var}(x_i) - \frac{\mathrm{Var}(x_i)}{n} = \frac{n-1}{n}\mathrm{Var}(x_i).
\end{aligned}$$

Consistency of the estimator

Given that fourth moments exist, the LLN tells us that

$$s^2 \xrightarrow{\text{p}} \frac{n-1}{n}\text{Var}(x_i) \xrightarrow{n\to\infty} \text{Var}(x_i)$$

and thus s^2 converges in probability to $\text{Var}(x_i)$ even though it is a biased estimator of this quantity; the bias goes to 0 as $n \to \infty$.

It is trivial to create an unbiased estimator here, which, of course, is also consistent:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Another derivation

For completeness, we see this result one other way:

$$\mathbb{E}[s^2] = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

and

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \xrightarrow{\text{p}} \mathbb{E}\left[x_i^2\right] \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{\text{p}} \mathbb{E}[x_i],$$

and a combination of these two results, with an application of the continuous mapping theorem applied to the latter, shows consistency.

Consistency and unbiasedness

It should be clear that an estimator could be unbiased and consistent (*e.g.*, , a simple sample mean) and biased and inconsistent (*e.g.*, the estimator 6).

A biased estimator can be consistent; the bias converges to 0 as n goes to infinity (*e.g.*, the preceding sample variance illustration, ratios of estimators).

An unbiased estimator can be inconsistent; the variance of the estimator may not shrink to 0 as n goes to infinity. See Davidson and Mackinnon (2004), page 97 for good examples.

Convergence in distribution

Let $\{b_n\}$ be a sequence of random variables/vectors with joint distribution $\{F_n\}$ and Z is a random variable/vector with distribution F. b_n *converges in distribution* to Z if

$$F_n(z) \overset{n\to\infty}{\longrightarrow} F(z)$$

for all continuity points z (*i.e.*, the subset of $\mathbb{R}^k$ such that $F(z)$ is continuous). This is often written as

$$b_n \overset{\mathrm{d}}{\longrightarrow} Z \text{ or } b_n \overset{\mathrm{A}}{\sim} F$$

and F is called the *limiting distribution* of b_n. F cannot be a function of n.

Example

This does *not* mean that b_n and Z have similar distributions. For example, let $X_n \sim \mathrm{N}(0, 1/n)$. Then,

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{2} & x = 0, \text{ and} \\ 1 & x > 0. \end{cases}$$

This function is not right continuous at 0, so it's technically not a CDF.

A better limiting distribution would be $F(x) = \mathbb{I}\{x \geq 0\}$. This implies that $\Pr(X_n = 0) = 1$; X_n is a *degenerate* random variable.

If $X_n \xrightarrow{\text{d}} c$, where c is a degenerate random variable, then $X_n \xrightarrow{\text{p}} c$. In our example, X_n converges in probability to 0.

Slutsky's theorem

Slustky's theorem states that, if $X_n \xrightarrow{\text{p}} X$ and $Y_n \xrightarrow{\text{d}} Y$, then

$$X_n + Y_n \xrightarrow{\text{d}} X + Y \text{ and}$$
$$X_n' Y_n \xrightarrow{\text{d}} X' Y$$

Central limit theorem

If $\{X_n\}$ are i.i.d. with mean μ and variance σ^2, then the *Lindeberg-Levy central limit theorem* states that

$$\frac{(\bar{x}_n - \mu)}{\sqrt{\mathrm{Var}(\bar{x})}} = \sqrt{n}\frac{(\bar{x}_n - \mu)}{\sigma} \xrightarrow{\mathrm{d}} \mathrm{N}(0, 1).$$

Rewriting this as

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2)$$

shows that the difference $\bar{x}_n - \mu$ gets smaller at a rate that is just offset by the increase in $\sqrt{n}$. We call this *root n convergence.*

Multivariate extension

The *Cramer-Wald device* says that if a scalar random variable $X_n'\lambda$ converges to $X'\lambda$, then the random vector X' converges to X.

A random vector with sample mean μ (a vector) and variance matrix Σ has

$$\sqrt{n}\left(\bar{X} - \mu\right) \xrightarrow{\text{d}} \mathrm{N}(0, \Sigma).$$

Linear model CLT

By the CLT, we have

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i\right) \xrightarrow{\text{d}} \mathrm{N}(0, \Sigma),$$

recalling that the sample average above converges to its expectation, which is assumed to be 0 with some variance-covariance matrix Σ.

Linear model CLT, continued

Now we have

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{\mathrm{d}} \mathbb{E}\left[xx'\right]^{-1} \times \mathrm{N}(0,\Sigma)$$
$$= \mathrm{N}\left(0, \mathbb{E}\left[xx'\right]^{-1} \Sigma \mathbb{E}\left[xx'\right]^{-1}\right)$$

using Slutsky's theorem.

The variance-covariance matrix $\mathbb{E}\left[xx'\right]^{-1} \Sigma \mathbb{E}\left[xx'\right]^{-1}$ is often called the *sandwich estimator* for the variance for obvious reasons.

Lingo

We say that the *limiting distribution* of $\sqrt{n}\left(\hat{\beta}-\beta\right)$ is $\mathrm{N}\left(0, \mathbb{E}\left[xx'\right]^{-1}\Sigma\mathbb{E}\left[xx'\right]^{-1}\right)$.

We say that the *asymptotic distribution* of $\left(\hat{\beta}-\beta\right)$ is

$$\mathrm{N}\left(0, \frac{1}{n}\mathbb{E}\left[xx'\right]^{-1}\Sigma\mathbb{E}\left[xx'\right]^{-1}\right).$$

Note that the asymptotic distribution can depend upon n, while the limiting distribution cannot.

The Σ matrix

To find Σ, we have

$$\begin{aligned}\Sigma &= \operatorname{Var}\left(x_i \epsilon_i\right) \\ &= \mathbb{E}\left[x_i \epsilon_i \epsilon_i x_i'\right] \\ &= \mathbb{E}\left[x_i x_i' \epsilon_i^2\right],\end{aligned}$$

using the fact that $\mathbb{E}\left[x_i \epsilon_i\right] = 0$.

If we assume that $\operatorname{Var}(\epsilon_i \mid x_i) = \sigma^2$, then this reduces to

$$\sigma^2 \mathbb{E}\left[x_i x_i'\right]$$

and the sandwich estimator becomes

$$\sigma^2 \left[\mathbb{E}\left[x_i x_i'\right]\right]^{-1}.$$