Econometrics I - Asymptotic Properties of Various Estimators

Ryan T. Godwin

University of Manitoba

So far our results apply for any finite sample size n. In more general models we often can't obtain exact results for estimators' properties (for example, models that are estimated via *maximum likelihood*. In these cases, we might instead consider the estimator's properties as $n \to \infty$, as a way of "approximating" results. Asymptotic properties of estimators are also of interest in their own right: inferential procedures should "work well" when we have lots of data. We have already seen one example of an asymptotic property: hypothesis tests that are "consistent".

Weak consistency. An estimator, $\hat{\boldsymbol{\theta}}$, for $\boldsymbol{\theta}$, is said to be (weakly) *consistent* if

$$\lim_{n\to\infty} \left\{ \Pr.\left[\left| \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right| < \boldsymbol{\epsilon} \right] \right\} = 1$$

for all $\boldsymbol{\epsilon} > \mathbf{0}$.

A sufficient condition for *weak* consistency to hold is the stronger mean-square consistency:

Mean-square consistency. An estimator, $\hat{\boldsymbol{\theta}}$, for $\boldsymbol{\theta}$, is said to be mean-square *consistent* if its bias and variance go to zero as n goes to infinity:

$$\begin{aligned} &\text{Bias}\left(\widehat{\boldsymbol{\theta}}_n\right) \to \mathbf{0};\ \text{as } n \to \infty, \\ &V\left(\widehat{\boldsymbol{\theta}}_n\right) \to 0;\ \text{as } n \to \infty. \end{aligned}$$

Mean-square consistency is often useful for checking consistency, since it is easier to prove than weak consistency.

If $\hat{\boldsymbol{\theta}}$ is weakly consistent for $\boldsymbol{\theta}$, we say that the "probability limit" of $\hat{\boldsymbol{\theta}}$ equals $\boldsymbol{\theta}$. We denote this by using the "plim" operator, and we write:

$$\text{plim}\left(\widehat{\boldsymbol{\theta}}_n\right) = \boldsymbol{\theta} \quad \text{or,} \quad \widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}$$

Loosely speaking, consistency means that, given an infinitely large sample of data, the estimator provides the true parameter value *exactly* (there is zero bias and variance).

Consistency of the sample average. Let $x_i \sim \left[\mu, \sigma^2\right]$ be a random i.i.d. variable.

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
E[\bar{x}] &= \frac{1}{n}\sum_{i=1}^{n} E\left(x_i\right) = \frac{1}{n}(n\mu) = \mu \qquad \text{(unbiased, for all } n) \\
\operatorname{var}[\bar{x}] &= \frac{1}{n^2}\operatorname{var}\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}\left(x_i\right) \\
&= \frac{1}{n^2}\left(n\sigma^2\right) = \sigma^2/n
\end{aligned}
$$

So, $\bar{x}$ is an unbiased estimator of μ, and $\lim_{n\to\infty}\{\operatorname{var}[\bar{x}]\} = 0$. This implies that $\bar{x}$ is both a mean-square consistent, and a weakly consistent estimator of μ.

Note:

- If an estimator is inconsistent, then it is a pretty useless estimator!
- There are many situations in which our LS estimator is inconsistent! For example, when:
  - there is an “omitted” variable that is relevant, and that is correlated to a variable included in the estimated model;
  - there is *simultaneous causality*;
  - a time series model includes lags of the dependent variable, and the errors are *autocorrelated*.

Slutsky's Theorem

Let $\text{plim}\left(\widehat{\boldsymbol{\theta}}_n\right) = \boldsymbol{c}$, and let $f(.)$ be any continuous function. Then, $\text{plim}\left[f\left(\widehat{\boldsymbol{\theta}}_n\right)\right] = f(\boldsymbol{c})$. For example:

$$\text{plim}\left(\frac{1}{\widehat{\theta}}\right) = \frac{1}{c}$$

where $\widehat{\theta}$ and c are scalars;

$$\text{plim}\left(e^{\hat{\boldsymbol{\theta}}}\right) = e^{\boldsymbol{c}}$$

where $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{c}$ are vectors;

$$\text{plim}\left(\hat{\Theta}^{-1}\right) = C^{-1}$$

where $\hat{\Theta}$ and C are matrices. Slutsky's Theorem is a very useful result: the "plim" operator can be used very flexibly.

Asymptotic Properties of LS Estimator

Consider LS estimator of β under our standard assumptions, in the "large n" asymptotic case.

- Can relax some assumptions:
  1. Don't need Normality assumption for the error term of our model.
  2. Columns of X can be random, just assume that $\{\boldsymbol{x}_i', \epsilon_i\}$ is a random and *independent* sequence; $i = 1, 2, 3, \ldots$.
  3. The above assumption implies that plim $\left[n^{-1} X' \epsilon\right] = \mathbf{0}$.
- Amend (extend) our assumption about X having full column rank. Assume instead that plim $\left[n^{-1} X' X\right] = Q$, where Q is a finite, positive-definite and symmetric $(k \times k)$ matrix that is *unobservable.*

Question: In words, what are we assuming about the elements of X, as n increases without limit?

Theorem: The LS estimator of $\boldsymbol{\beta}$ is weakly consistent.

Proof:

$$\begin{aligned}\boldsymbol{b} = \left(X'X\right)^{-1} X'\boldsymbol{y} &= \left(X'X\right)^{-1} X'(X\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\ &= \boldsymbol{\beta} + \left(X'X\right) X'\boldsymbol{\epsilon} \\ &= \boldsymbol{\beta} + \left[\frac{1}{n}\left(X'X\right)\right]^{-1} \left[\frac{1}{n}X'\boldsymbol{\epsilon}\right].\end{aligned}$$

If we now apply Slutsky's Theorem repeatedly, we have:

$$\text{plim}(\boldsymbol{b}) = \boldsymbol{\beta} + Q^{-1}\mathbf{0} = \boldsymbol{\beta}$$

We can also show that s^2 is a consistent estimator for σ^2. There are at least two ways to do this (each uses different assumptions). First, assume the errors are Normally distributed, and get a strong result. We can also relax this assumption and get a weaker result.

Theorem:

If the regression model errors are Normally distributed, then s^2 is a *mean-square* consistent estimator for σ^2.
Proof: If the errors are Normal, then we know that

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2_{(n-k)}$$

The mean and variance of a χ^2 distributed random variable are:

$$E\left[\chi^2_{(n-k)}\right] = (n-k)$$

$$\text{var}\left[\chi^2_{(n-k)}\right] = 2(n-k)$$

So,

$$E\left(s^2\right) = \frac{\sigma^2 E\left[\chi^2_{(n-k)}\right]}{n-k} = \sigma^2 \quad ; \qquad \text{unbiased}$$

and

$$\text{var}\left[\frac{(n-k)s^2}{\sigma^2}\right] = 2(n-k)$$

$$\left[\frac{(n-k)^2}{\sigma^4}\right]\text{var}\left(s^2\right) = 2(n-k)$$

$$\text{var}\left(s^2\right) = 2\sigma^4/(n-k)$$

So, $\text{var}\left(s^2\right) \to 0$, as $n \to \infty$, and the estimator is unbiased. This implies that s^2 is a mean-square consistent estimator for σ^2. (This implies that it is also a weakly consistent estimator.)

- With the addition of the (relatively) strong assumption of Normally distributed errors, we get the (relatively) strong result.
- Note that $\hat{\sigma}^2 = (e'e)/n$ is also a consistent estimator, even though it is biased.

What can we say if we relax the assumption of Normality? We need a preliminary result to help us (Khintchine's theorem; or the Weak Law of Large Numbers).

Khinchin's Theorem; Weak Law of Large Numbers (WLLN)

Suppose that $\{x_i\}_{i=1}^{n}$ is a sequence of random variables that are uncorrelated, and all drawn from the same distribution with a finite mean, μ, and a finite variance, σ^2. Then, $\text{plim}(\bar{x}) = \mu$. Khinchin's theorem says that a sample average of i.i.d. variables is at least *weakly* consistent. We can use this result to establish the consistency of s^2.

Weak consistency of s^2. In our regression model, s^2 is a weakly consistent estimator for σ^2. (Notice that this also means that $\hat{\sigma}^2$ is a weakly consistent estimator, so start with the latter estimator.)
Proof:

$$\begin{aligned}
\hat{\sigma}^2 &= \left(\frac{e'e}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \\
&= \frac{1}{n}(M\epsilon)'(M\epsilon) = \frac{1}{n}\epsilon' M\epsilon \\
&= \frac{1}{n}\left[\epsilon'\epsilon - \epsilon' X\left(X'X\right)^{-1} X'\epsilon\right] \\
&= \left[\left(\frac{1}{n}\epsilon'\epsilon\right) - \left(\frac{1}{n}\epsilon' X\right)\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\epsilon\right)\right]
\end{aligned}$$

So, $\text{plim}\left(\hat{\sigma}^2\right) = \text{plim}\left(\frac{1}{n}\boldsymbol{\epsilon}'\boldsymbol{\epsilon}\right) - \mathbf{0}'Q^{-1}\mathbf{0} = \text{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2\right]$ (if the errors are not autocorrelated, neither are the squared values). Also, $E\left[\epsilon_i^2\right] = \text{var}.\left(\epsilon_i\right) = \sigma^2$. By Khintchine's Theorem, we immediately have the result:

$$\text{plim}\left(\hat{\sigma}^2\right) = \sigma^2$$

and

$$\text{plim}\left(s^2\right) = \sigma^2$$

Relaxing the assumption of Normally distributed errors led to a weaker result for the consistent estimation of the error variance.

Asymptotic efficiency

Suppose we want to compare the (large n) asymptotic behaviour of our LS estimators with those of other potential estimators. These other estimators will presumably also be consistent. This means that in each case the sampling distributions of the estimators collapse to a "spike", located exactly at the true parameter values. So, how can we compare such estimators when n is very large: aren't they indistinguishable? If the limiting density of any consistent estimator is a degenerate "spike", it will have zero variance, in the limit. Can we still compare large-sample variances of consistent estimators? In other words, is it meaningful to think about the concept of asymptotic efficiency?

The key to asymptotic efficiency is to "control" for the fact that the distribution of any consistent estimator is "collapsing", as $n \to \infty$.

The rate at which the distribution collapses is crucially important. This is probably best understood by considering an example.

Example

Let $\{x_i\}_{i=1}^n$ be a random sample from a population with mean and variance $\left[\mu, \sigma^2\right]$. We know from a previous example that:

$$E[\bar{x}] = \mu; \quad \text{var}[\bar{x}] = \sigma^2/n$$

Observe how $\text{var}[\bar{x}] \to 0$ as $n \to \infty$ (the sampling distribution collapses to a "spike" at the true parameter value). Now, construct: $y = \sqrt{n}(\bar{x} - \mu)$. Note that:

$$E(y) = \sqrt{n}(E(\bar{x}) - \mu) = 0$$

and

$$\text{var}[y] = (\sqrt{n})^2 \, \text{var}(\bar{x} - \mu) = n \, \text{var}(\bar{x}) = \sigma^2$$

The scaling we've used results in a finite, non-zero, variance. $E(y) = 0$, and $\text{var}[y] = \sigma^2$; *unchanged* as $n \to \infty$. So, $y = \sqrt{n}(\bar{x} - \mu)$ has a well-defined "limiting" (asymptotic) distribution. The asymptotic mean of y is zero, and the asymptotic variance of y is σ^2. Why did we scale by $\sqrt{n}$, and not (say), by n itself?

In fact, because we had independent x_i's (random sampling), we have the additional result that $y = \sqrt{n}(\bar{x} - \mu) \xrightarrow{d} N\left[0, \sigma^2\right]$, the Lindeberg-Lévy Central Limit Theorem.

Definition

Let $\hat{\theta}$ and $\tilde{\theta}$ be two consistent estimator of θ; and suppose that

$$\sqrt{n}(\hat{\theta} - \theta) \overset{d}{\to} \left[0, \sigma^2\right], \text{ and } \sqrt{n}(\tilde{\theta} - \theta) \overset{d}{\to} \left[0, \varphi^2\right].$$

Then $\hat{\theta}$ is "asymptotically efficient" relative to $\tilde{\theta}$ if $\sigma^2 < \varphi^2$. In the case where $\boldsymbol{\theta}$ is a vector, $\widehat{\boldsymbol{\theta}}$ is "asymptotically efficient" relative to $\widetilde{\boldsymbol{\theta}}$ if $\Delta = \operatorname{asy}.V(\widetilde{\boldsymbol{\theta}}) - \operatorname{asy}.V(\widehat{\boldsymbol{\theta}})$ is positive definite.

Asymptotic Distribution of the LS Estimator

Let's consider the full asymptotic distribution of the LS estimator, $\boldsymbol{b}$, for $\boldsymbol{\beta}$ in our linear regression model. We'll actually have to consider the behaviour of $\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})$:

$$\begin{aligned}\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta}) &= \sqrt{n}\left[\left(X'X\right)^{-1}X'\boldsymbol{\epsilon}\right] \\ &= \left[\frac{1}{n}\left(X'X\right)\right]^{-1}\left(\frac{1}{\sqrt{n}}X'\boldsymbol{\epsilon}\right)\end{aligned}$$

It can be shown, by the Lindeberg-Feller Central Limit Theorem, that

$$\left(\frac{1}{\sqrt{n}}X'\boldsymbol{\epsilon}\right) \xrightarrow{d} N\left[0, \sigma^2 Q\right]$$

where $Q = \text{plim}\left[\frac{1}{n}\left(X'X\right)\right]$. So, the asymptotic covariance matrix of $\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})$ is

$$\text{plim}\left[\frac{1}{n}\left(X'X\right)\right]^{-1}\left(\sigma^2 Q\right)\text{plim}\left[\frac{1}{n}\left(X'X\right)\right]^{-1} = \sigma^2 Q^{-1}.$$

In full, the asymptotic distribution of $\boldsymbol{b}$ is correctly stated by saying that:

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

The asymptotic covariance matrix is unobservable, for two reasons:

1. σ^2 is typically unknown.
2. Q is unobservable.

- We can estimate σ^2 consistently, using s^2.
- To estimate $\sigma^2 Q^{-1}$ consistently, we can use $ns^2 \left(X'X\right)^{-1}$:

$$\operatorname{plim}\left[ns^2\left(X'X\right)^{-1}\right] = \operatorname{plim}\left(s^2\right)\operatorname{plim}\left[\frac{1}{n}\left(X'X\right)\right]^{-1} = \sigma^2 Q^{-1}$$

The square roots of the diagonal elements of $ns^2\left(X'X\right)^{-1}$ are the asymptotic standard errors for the elements of $\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})$. Loosely speaking, the asymptotic covariance matrix for $\boldsymbol{b}$ itself is $s^2\left(X'X\right)^{-1}$; and the square roots of the diagonal elements of this matrix are the asymptotic standard errors for the b_i's themselves.