

LIMIT OF A SEQUENCE

Suppose that $\{a_1, a_2, \dots, a_n\}$ constitutes a sequence of real numbers. If there exists a real number d such that for every real number ϵ however small there exists an integer $N(\epsilon)$ such that for all $n > N(\epsilon)$ $|a_n - d| < \epsilon$, then we say that d is the limit of the sequence $\{a_n\}$ and write $\lim_{n \rightarrow \infty} \{a_n\} = d$.

LIMIT OF A FUNCTION

If the function $f(x)$ has the limit A at the point x_0 if for every $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that $|f(x) - A| < \epsilon$ whenever $0 < |x - x_0| < \delta(\epsilon)$.

CONVERGENCE IN DISTRIBUTION

Given a sequence of random variable X_n whose CDF is $F_n(x)$ and a cumulative distribution function $F(x)$ which corresponds to random variable X , we say that X_n converges in distribution to X and write

$$X_n \xrightarrow{d} X \quad \text{if}$$

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \text{ at all points } x \text{ at which } F(x) \text{ is continuous.}$$

CONVERGENCE IN PROBABILITY

The sequence of random variables X_n converges in probability to the real number x if $\lim_{n \rightarrow \infty} P[|X_n - x| > \epsilon] = 0$ for each $\epsilon > 0$.

Thus, it becomes less and less likely that the random variable $(X_n - x)$ lies outside the interval $(\pm \epsilon)$.

The following statements are equivalent:

- (a) $\lim_{n \rightarrow \infty} P(|X_n - x| < \epsilon) = 1, \epsilon > 0.$
 - (b) Given $\epsilon > 0$ and $\delta > 0 \exists N(\epsilon, \delta)$ such that $P(|X_n - x| > \epsilon) < \delta$ for all $n > N(\epsilon, \delta).$
 - (c) $P(|X_n - x| < \epsilon) > 1 - \delta$ for all $n > N.$
- That is, $P(|X_{N+1} - x| < \epsilon) > 1 - \delta, P(|X_{N-1} - x| < \epsilon) > 1 - \delta$ etc.

Convergence in probability is indicated by

$X_n \xrightarrow{P} x$ or $\boxed{\text{plim } X_n = x}$.

A familiar example of convergence in probability is: $\text{plim } \bar{X}_n = \mu.$

Remember that $E(\bar{X}_n) = \mu$ and $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$

As the sample size (n) increases, the variance of the sampling distribution of the mean becomes smaller and concentrated around the mean (μ) . Hence, the probability mass gets concentrated around the mean (μ) . Hence, there is a large enough n such that $\Pr[\bar{X}_n \in (\mu \pm \epsilon)] > 1 - \delta.$

MEAN SQUARED ERROR CONVERGENCE

A random variable X_n converges in the quadratic mean to x if $E[(X_n - x)^2]$ exists and

$$\lim_{n \rightarrow \infty} E[(X_n - x)^2] = 0.$$

Again, the sample mean converges in the mean-squares to the population mean.

$$E[(\bar{X}_n - \mu)^2] = \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

and $\lim_{n \rightarrow \infty} \text{Var}(\frac{\sigma^2}{n}) \rightarrow 0.$

ALMOST SURE CONVERGENCE:

The sequence of random variables X_n is said to converge almost surely to the real number x and is written as

$$P[\lim_{n \rightarrow \infty} X_n = x] = 1 \quad \text{a.s.}$$

This means that given ϵ and $\delta > 0$, there exist an integer N such that

$$Pr[|X_{N+1} - x| < \epsilon, |X_{N+2} - x| < \epsilon, \dots] > 1 - \delta.$$

That is, the joint probability that these values stay arbitrarily close to x can be made arbitrarily close to 1.

CONSISTENCY

SIMPLE CONSISTENCY:

Let $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n$ be a sequence of estimators of the parameter θ . Then $\hat{\theta}_n$ is a simple consistent estimator of θ if for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr [|\hat{\theta}_n - \theta| < \epsilon] = 1.$$

This implies that $\text{plim } \hat{\theta}_n = \theta$.

SQUARED-ERROR CONSISTENCY:

The sequence $(\hat{\theta}_n)$ is a squared error consistent estimator of θ if $\lim_{n \rightarrow \infty} E [(\hat{\theta}_n - \theta)^2] = 0$.

STRONG CONSISTENCY:

$\hat{\theta}_n$ is a strongly consistent estimator of θ if $P [\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta] = 1$

Note that ~~some~~ strong consistency requires almost sure convergence

UNBIASEDNESS & CONSISTENCY

consider a random sample $\{x_1, x_2, \dots, x_n\}$ drawn from the population with $E(x) = \mu$. Let $\bar{x}_1 = \bar{\mu}$ be an estimator of μ . clearly, $E(\bar{x}_1) = \mu$ so that $\bar{\mu}$ is an unbiased estimator. But \bar{x}_1 does not converge in probability to μ no matter how large the sample size becomes. $\bar{\mu}$ is not a consistent estimator.

ASYMPTOTIC UNBIASEDNESS

One often characterizes an estimator $\hat{\theta}_n$ as asymptotically unbiased if $\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta$. But this definition is not useful if $E(\hat{\theta}_n)$ does not exist.

Alternatively, $\hat{\theta}_n$ is an asymptotically unbiased estimator of θ if the mean of the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ is zero.

Asymptotic unbiasedness implies

$$AE(\hat{\theta}_n) = \theta$$

Similarly, the asymptotic variance can be defined as either

$$\lim_{n \rightarrow \infty} E \left[\sqrt{n} (\hat{\theta}_n - \theta) \right]^2 \text{ or } AE \left[\left\{ \sqrt{n} (\hat{\theta}_n - \theta) \right\}^2 \right]$$

The first is the limit of the variance. The other is the variance of the limiting distribution. Note that in defining asymptotic variance we consider $\sqrt{n}(\hat{\theta}_n - \theta)$ so that variances do not tend to 0 and n increases.

ASYMPTOTIC UNBIASEDNESS & CONSISTENCY

Let $x = (x_1, x_2, \dots, x_n)$ be a random sample drawn from $N(\theta, 1)$. Let $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n x_i$ be an estimator of θ .

$$E[\hat{\theta}_n] = \frac{1}{2}\theta + \frac{n-1}{2n}\theta$$

$$= \frac{n\theta + (n-1)\theta}{2n} = \left(1 - \frac{1}{2n}\right)\theta$$

$\lim_{n \rightarrow \infty} E[\hat{\theta}_n] = \theta$. Thus $\hat{\theta}_n$ is an asymptotically unbiased estimator

unbiased estimator

But

$$\text{plim } \hat{\theta}_n = \text{plim} \left(\frac{1}{2}x_1 \right) + \text{plim} \frac{(n-1)\bar{x}_{n-1}}{2n}$$

$$= \frac{1}{2}x_1 + \frac{1}{2}\theta \neq \theta$$

Hence, asymptotic consistency does not ensure unbiasedness.

STOCHASTIC REGRESSORS:

Consider the 2 variable regression

$$Y_i = \alpha + \beta X_i + u_i$$

and OLS estimator

$$\hat{\beta} = \frac{\sum (X_i - \bar{x})(Y_i - \bar{y})}{\sum (X_i - \bar{x})^2} = \beta + \frac{\sum (X_i - \bar{x})(u_i - \bar{u})}{\sum (X_i - \bar{x})^2}$$

$$= \beta + \frac{\sum x_i u_i}{\sum x_i^2} \quad \text{where } x_i \equiv X_i - \bar{x}$$

Note that $E(\hat{\beta}) = \beta$ when

$$E \left[\frac{x_1 u_1 + x_2 u_2 + \dots + x_n u_n}{\sum x_i^2} \right] = 0$$

When the x_i are non random, this expectation is

$$\frac{x_1}{\sum x_i^2} E(u_1) + \frac{x_2}{\sum x_i^2} E(u_2) + \dots + \frac{x_n}{\sum x_i^2} E(u_n) = 0$$

Even when the x_i are random so that $\frac{x_i}{\sum x_i^2}$ is also random, so long as they are independent of the u_i , $E(\hat{\beta}) = \beta + \sum_i E\left(\frac{x_i}{\sum x_i^2}\right) E(u_i) = \beta$

because $E(u_i) = 0$ no matter what

$E\left[\frac{x_i}{\sum x_i^2}\right]$ is

Even when x and u are not independent,

$$\text{plim } \frac{1}{n} \sum x_i u_i = 0$$

$$\text{if } \text{plim } \hat{\beta} = \beta + \text{plim } \frac{\sum x_i u_i}{\sum x_i^2} = \beta + \frac{\text{plim } \frac{1}{n} \sum x_i u_i}{\text{plim } \frac{1}{n} \sum x_i^2} = \beta$$

Thus $\hat{\beta}$ would be a consistent estimator of β

Now suppose that the true model is

$$Y_i = \alpha + \beta x_i + u_i$$

but we can observe only $Z_i = x_i + v_i$.

Thus
$$Y_i = \alpha + \beta (Z_i - v_i) + u_i$$

$$Y_i = \alpha + \beta Z_i + (u_i - \beta v_i)$$

Define

$$W_i = u_i - \beta v_i$$

Then

$$Y_i = \alpha + \beta Z_i + W_i$$

$$\hat{\beta} = \frac{\sum (z_i - \bar{z})(y_i - \bar{y})}{\sum (z_i - \bar{z})^2} = \beta + \frac{\sum (z_i - \bar{z})(w_i - \bar{w})}{\sum (z_i - \bar{z})^2}$$

Note that $z_i - \bar{z} = (x_i - \bar{x}) + (v_i - \bar{v})$
and $w_i - \bar{w} = (u_i - \bar{u}) - \beta(v_i - \bar{v})$

Thus
$$\frac{1}{n} \sum (z_i - \bar{z})(w_i - \bar{w})$$

$$= \text{Cov}(x, u) - \beta \text{Cov}(x, v) + \text{Cov}(u, v) - \beta \text{Var}(v)$$

$$\frac{1}{n} \sum (z_i - \bar{z})^2 = \text{Var}(z) = \text{Var}(x) + \text{Var}(v) + 2\text{Cov}(x, v)$$

Assume that
$$\begin{aligned} \text{plim Cov}(x, u) &= \\ \text{plim Cov}(u, v) &= \\ \text{plim Cov}(x, v) &= 0 \end{aligned}$$

Thus
$$\text{plim } \hat{\beta} = \beta + \frac{\text{plim Cov}(z, w)}{\text{plim Var}(z)}$$

$$= \beta - \frac{\beta \sigma_v^2}{\sigma_x^2 + \sigma_v^2} = \beta \left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_v^2} \right)$$

$< \beta$

Assume that we have the "true" regression model

$$y = \tilde{x}\beta + u.$$

But the matrix of observed values of the explanator variables is $X = \tilde{X} + V \Leftrightarrow \tilde{X} = X - V$

$$\text{Thus, } y = (X - V)\beta + u \\ \Rightarrow y = X\beta + (u - V\beta).$$

If we look at the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$,
 $\hat{\beta} = \beta + (X'X)^{-1}X'(u - V\beta).$

We may assume:

1) that the measurement errors V are uncorrelated in the limit with the true values \tilde{X} .
Thus, $\text{plim} \frac{\tilde{X}'V}{n} = 0$

$$\text{Hence, } \text{plim} \left(\frac{1}{n} X'X \right) = \text{plim} \left(\frac{1}{n} \tilde{X}'\tilde{X} \right) + \text{plim} \left(\frac{1}{n} V'V \right) \\ = \Sigma + \Omega$$

$$\text{where } \Sigma = \text{plim} \left(\frac{1}{n} \tilde{X}'\tilde{X} \right) \text{ \& } \Omega = \text{plim} \left(\frac{1}{n} V'V \right)$$

2) These measurement errors (V) are uncorrelated in the limit with the disturbance term u . Thus
 $\text{plim} \left(\frac{1}{n} V'u \right) = 0$

But by specification \tilde{X} & u are independent.
Hence $\text{plim} \left(\frac{1}{n} \tilde{X}'u \right) = 0$ also holds

$$\text{then } \text{plim}(\hat{\beta}) = \beta + \text{plim} \left(\frac{X'X}{n} \right)^{-1} \text{plim} \left(\frac{X'u}{n} \right) \\ - \text{plim} \left(\frac{X'X}{n} \right)^{-1} \text{plim} \left(\frac{X'V}{n} \right) \beta$$

$$\text{Now } \text{plim} \left(\frac{X'V}{n} \right) \beta = \text{plim} \left(\frac{\tilde{X}'V}{n} + \frac{V'V}{n} \right) \beta \\ = \text{plim} \left(\frac{V'V}{n} \right) \beta = \Omega \beta$$

Therefore $\text{plim } \hat{\beta} = \beta + (\Sigma + \Omega)^{-1} \Omega \beta$

Thus, OLS estimators are not consistent. The inconsistency arises out of interdependence between X and the composite error $(u - V\beta)$.

Example:

$$y_i = \alpha + \beta \tilde{x}_i + u_i$$

$$x_i = \tilde{x}_i + v_i$$

Thus $y_i = \alpha + \beta x_i + (u_i - \beta v_i)$

$$\Sigma = \text{plim } \frac{1}{n} (\tilde{X} \tilde{X}') = \text{plim } \begin{bmatrix} 1 & \frac{1}{n} \sum \tilde{x}_i \\ \frac{1}{n} \sum \tilde{x}_i & \frac{1}{n} \sum \tilde{x}_i^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \mu \\ \mu & \mu^2 + \sigma^2 \end{bmatrix}$$

Here μ & σ^2 are the mean and the variance of \tilde{x}_i .

$$\Omega = \text{plim } \left(\frac{1}{n} V'V \right) = \text{plim } \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{n} \sum v_i^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}$$

$$\Sigma + \Omega = \begin{bmatrix} 1 & \mu \\ \mu & \mu^2 + \sigma^2 + \sigma_v^2 \end{bmatrix}$$

$$\left[\Sigma + \Omega \right]^{-1} = \frac{\begin{bmatrix} \mu^2 + \sigma^2 + \sigma_v^2 & -\mu \\ -\mu & 1 \end{bmatrix}}{\sigma^2 + \sigma_v^2}$$

$$\left[\Sigma + \Omega \right]^{-1} \Omega = \frac{1}{\sigma^2 + \sigma_v^2} \begin{bmatrix} 0 & -\mu \sigma_v^2 \\ 0 & +\sigma_v^2 \end{bmatrix}$$

$$\text{plim} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - \frac{1}{\sigma^2 + \sigma_v^2} \begin{bmatrix} 0 - \mu \sigma_v^2 \\ 0 - \sigma_v^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$= \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - \frac{1}{\sigma^2 + \sigma_v^2} \begin{bmatrix} -\mu \sigma_v^2 \beta \\ + \sigma_v^2 \beta \end{bmatrix}$$

$$\text{plim} \hat{\alpha} = \alpha + \mu \frac{\sigma_v^2}{\sigma^2 + \sigma_v^2} \beta$$

$$\text{plim} \hat{\beta} = \beta - \frac{\sigma_v^2}{\sigma^2 + \sigma_v^2} \beta$$

$$= \underline{\underline{\beta \frac{\sigma^2}{\sigma^2 + \sigma_v^2}}}$$

INSTRUMENTAL VARIABLE ESTIMATOR

The problem of inconsistency arises because the measured values x are correlated with the composite error $(u - v\beta)$.

Now assume that we can replace x s by a matrix Z where the Z s are correlated with u or v but not correlated with u or v .

~~Now~~ consider the estimator

$$\beta_{IV} = (Z'X)^{-1} Z'Y$$

$$= (Z'X)^{-1} Z'(X\beta + u - v\beta)$$

$$= \beta + (Z'X)^{-1} Z'u - (Z'X)^{-1} Z'v\beta$$

$$\text{plim} \beta_{IV} = \beta + \text{plim} \left(\frac{Z'X}{n} \right)^{-1} \text{plim} \left(\frac{Z'u}{n} \right)$$

$$= \beta + \Sigma_{xz}^{-1} 0 - \Sigma_{zx}^{-1} 0 \cdot \beta = \beta$$

Consider again the regression

$$y_i = \alpha + \beta x_i + (u - \beta v_i)$$

Suppose that there are n observations and n is even. Then define a variable z which takes the value 1 if x_i is above the median value and -1 if x_i is less than the median value of x . Now consider the matrix

$$Z = \begin{bmatrix} 1 & -1 \\ 1 & +1 \\ \vdots & \vdots \\ +1 & -1 \end{bmatrix}$$

with this definition,

$$\begin{bmatrix} \alpha_{IV} \\ \beta_{IV} \end{bmatrix} = (Z'X)^{-1} Z'y$$

$$Z'X = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ -1 & +1 & \dots & +1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ 0 & \frac{n}{2}(\bar{x}_{II} - \bar{x}_{I}) \end{bmatrix}$$

where \bar{x}_I = mean of values of x less than median
 \bar{x}_{II} = " " " " " x greater " " "

$$Z'y = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ -1 & +1 & \dots & +1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ \frac{n}{2}(\bar{y}_{II} - \bar{y}_I) \end{bmatrix}$$

$$(Z'X) \begin{pmatrix} \alpha_{IV} \\ \beta_{IV} \end{pmatrix} = Z'y$$

$$\Rightarrow \begin{bmatrix} n & n\bar{x} \\ 0 & \frac{n}{2}(\bar{x}_U - \bar{x}_L) \end{bmatrix} \begin{bmatrix} \alpha_{IV} \\ \beta_{IV} \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ \frac{n}{2}(\bar{y}_U - \bar{y}_L) \end{bmatrix}$$

$$n \alpha_{IV} + n \beta_{IV} \bar{x} = n \bar{y}$$

$$\frac{n}{2} (\bar{x}_U - \bar{x}_L) \beta_{IV} = \frac{n}{2} (\bar{y}_U - \bar{y}_L)$$

$$\Rightarrow \beta_{IV} = \frac{\bar{y}_U - \bar{y}_L}{\bar{x}_U - \bar{x}_L}$$

and $\alpha_{IV} = \bar{y} - \beta_{IV} \bar{x}$.

If n is odd, one should discard the central observation before applying this procedure. Under fairly general conditions the IV estimator (due to Wald) is consistent. But it usually has large standard errors.