Econometrics

Chapter 5. Multiple Regression: Asymptotic Properties

In Choi

Sogang University

Consistency

- Under the assumptions given earlier, OLS is BLUE.
- In other cases it is not always possible to find unbiased estimators.
- In those cases, we may settle for estimators that are consistent.
- Consistency means that the distribution of the estimator collapses to the true parameter value as $n \longrightarrow \infty$.

Consistency

Example

Autoregressive model of order 1 (AR(1) model)

$$y_t = \alpha y_{t-1} + u_t, u_t \sim iid(0, \sigma^2), \; |\alpha| < 1, \; t = 2, ..., n.$$

We cannot have $E(u_t \mid y_1, ..., y_n) = 0$ because $y_t = u_t + \alpha u_{t-1} + \alpha^2 u_{t-2} + ...$ Thus, the conditions for the Gauss-Markov theorem are not satisfied. The OLS estimator of α, $\hat{\alpha} = \frac{\sum_{t=2}^{n} y_t y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2}$, is biased. However, we can show that it is consistent.

Consistency of OLS

- Under the assumptions given earlier, the OLS estimator is consistent.
- Consistency can be proved for the simple regression case in a manner similar to the proof of unbiasedness.

Convergence in probability

- Let X be a random variable and $\{X_n\}$ a sequence of random varialbes. If for $\varepsilon > 0$

$$\lim_{n \to \infty} P\left(|X_n - X| > \varepsilon\right) = 0,$$

X_n is said to converge in probability to X, written

$$X_n \xrightarrow{p} X.$$

X is known as the probability limit of X_n, written

$$X = plim_{n \to \infty} X_n.$$

Convergence in probability

Chebyshev's inequality For $\varepsilon > 0$,

$$P[|X_n - X| > \varepsilon] \leq E(X_n - X)^2/\varepsilon^2.$$

Proof. Consider an indicator function that take value 1 if $|X_n - X| > \varepsilon$ and zero otherwise. Write this as $1\{|X_n - X| > \varepsilon\}$. Then,

$$P[|X_n - X| > \varepsilon] = E1\{|X_n - X| > \varepsilon\}.$$

If $|X_n - X| > \varepsilon$, $\frac{|X_n - X|^2}{\varepsilon^2} > 1$ and $1\{|X_n - X| > \varepsilon\} < \frac{|X_n - X|^2}{\varepsilon^2}$. If $|X_n - X| \leq \varepsilon$, $0 = 1\{|X_n - X| > \varepsilon\} \leq \frac{|X_n - X|^2}{\varepsilon^2}$ (≥ 0). Thus,

$$1\{|X_n - X| > \varepsilon\} \leq \frac{|X_n - X|^2}{\varepsilon^2} \text{ with probability 1,}$$

which gives $E1\{|X_n - X| > \varepsilon\} \leq \frac{E|X_n - X|^2}{\varepsilon^2}$.

Convergence in probability

Example

The weak law of large numbers

Let $\{X_i, i \geq 1\}$ be a sequence of iid r.v.s with mean μ and variance σ^2. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu \text{ as } n \to \infty,$$

because for $\varepsilon > 0$

$$P[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon] \leq E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right]^2 / \varepsilon^2 = \frac{\sigma^2}{\varepsilon^2 n} \to 0$$

as $n \to \infty$.

Convergence in probability

Some useful results regarding stochastic convergence are:

1. $X_n \xrightarrow{p} X$ and $g(\cdot)$ is a continuous function

$$\Rightarrow g(X_n) \xrightarrow{p} g(X).$$

Example

Let

$$X_n = \begin{cases} 1 & \text{with probability } \frac{1}{n} \\ 0 & \text{with probability } 1 - \frac{1}{n} \end{cases}.$$

Obviously, $X_n \xrightarrow{p} 0$. Let $g(x) = x + 1$. Then, $g(X_n) \xrightarrow{p} g(0) = 1$.

Convergence in probability

2. Suppose that $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$. Then

$$X_n + Y_n \xrightarrow{p} X + Y$$

$$X_n Y_n \xrightarrow{p} XY$$

$$\frac{Y_n}{X_n} \xrightarrow{p} \frac{Y}{X} \text{ when } X \neq 0 \text{ with probability 1.}$$

Consistency of OLS in simple linear regressions

- Consider the simple linear regression model

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

for which we assume that $\{x_t\}$ is a sequence of constants. Recall that

$$\hat{\beta}_1 - \beta_1 = \sum_{t=1}^{n} w_t u_t,$$

where $w_t = \frac{x_t - \bar{x}}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$.

Consistency of OLS in simple linear regressions

- Using Chebyshev's inequality, we find

$$
\begin{aligned}
P[|\hat{\beta}_1 - \beta_1| &> \varepsilon] \leq E(\hat{\beta}_1 - \beta_1)^2/\varepsilon^2 \\
&= \frac{\sigma^2}{\varepsilon^2 \sum_{t=1}^{n}(x_t - \bar{x})^2}.
\end{aligned}
$$

Thus, if $\sum_{t=1}^{n}(x_t - \bar{x})^2 \longrightarrow \infty$ as $n \longrightarrow \infty$, $\hat{\beta}_1 \xrightarrow{p} \beta_1$ as as $n \longrightarrow \infty$. Or if the variance of $\hat{\beta}_1$ goes to zero as $n \longrightarrow \infty$, the OLS estimator is consistent.

Consistency of OLS in simple linear regressions

- An alternative way of studying consistency of OLS: Write

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{t=1}^{n} (x_t - \bar{x})\, u_t / n}{\sum_{t=1}^{n} (x_t - \bar{x})^2 / n}$$

and assume $\sum_{t=1}^{n} (x_t - \bar{x})^2 / n \longrightarrow M \; (> 0)$. Then,

$$\begin{aligned} P\left[\left| \frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})\, u_t \right| \right. & > \left. \varepsilon \right] \leq \frac{1}{n^2} E\left(\sum_{t=1}^{n} (x_t - \bar{x})\, u_t\right)^2 / \varepsilon^2 \\ & = \frac{\sigma^2 \sum_{t=1}^{n} (x_t - \bar{x})^2}{n^2 \varepsilon^2} \rightarrow 0. \end{aligned}$$

Thus, $\hat{\beta}_1 - \beta_1 \xrightarrow{p} \frac{0}{M} = 0$.

A useful fact If $a_n \xrightarrow{p} a$ and $b_n \xrightarrow{p} b$ as $n \rightarrow \infty$, $\frac{a_n}{b_n} \xrightarrow{p} \frac{a}{b}$ as $n \rightarrow \infty$.

Consistency of OLS in simple linear regressions

- If $\{x_t\}$ is a sequence of r.v.'s, assume
  1. $\sum_{t=1}^{n}(x_t - \bar{x})^2/n \xrightarrow{p} M \ (>0)$
  2. $\bar{x} \xrightarrow{p} L$
  3. $Ex_t^2 < c \ (\in R)$ for all t
  4. $$E\left(u_t | all\ x's\right) = 0 \text{ for all } t.$$

  and
  $$\begin{cases} Var\left(u_t | all\ x's\right) = \sigma^2 \text{ for all } t = 1, 2, \cdots, n \\ Cov\left(u_t,\ u_s | all\ x's\right) = 0 \text{ for all } s \neq t. \end{cases},$$

Consistency of OLS in simple linear regressions

- Write

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{t=1}^{n} (x_t - \bar{x}) u_t / n}{\sum_{t=1}^{n} (x_t - \bar{x})^2 / n}.$$

The denominator converges to M in probability. For the nominator, we have

$$\begin{aligned} P[\left| \frac{1}{n} \sum_{t=1}^{n} x_t u_t \right| &> \varepsilon] \leq \frac{1}{n^2} EE \left\{ (\sum_{t=1}^{n} x_t u_t)^2 | all\ x's \right\} / \varepsilon^2 \\ &= \frac{\sigma^2 \sum_{t=1}^{n} E x_t^2}{n^2 \varepsilon^2} < \frac{\sigma^2 c}{n \varepsilon^2} \to 0 \end{aligned}$$

due to Assumptions 3 and 4, and $\bar{x} \frac{1}{n} \sum_{t=1}^{n} u_t \xrightarrow{p} 0$, which imply that the nominator converges to 0 in probability. Thus, $\hat{\beta}_1 - \beta_1 \xrightarrow{p} 0$.

Consistency of OLS in simple linear regressions

- Using this method,we can show consistency of the OLS estimator for the AR(1) model. Write

$$\hat{\alpha} - \alpha = \frac{\sum_{t=2}^{n} y_{t-1} u_t}{\sum_{t=2}^{n} y_{t-1}^2}.$$

  Then, if $|\alpha| < 1$, we can show that $\frac{1}{n}\sum_{t=2}^{n} y_{t-1}^2 \xrightarrow{p} \frac{\sigma^2}{1-\alpha^2}$ and that $\frac{1}{n}\sum_{t=2}^{n} y_{t-1} u_t \xrightarrow{p} 0$.
- For unbiasedness, we assumed a zero conditional mean $E(u_t | all\ x) = 0$. This implies $E(u_t f(all\ x)) = 0$, where $f(\cdot)$ is any function.

Consistency of OLS in simple linear regressions

- Consistency can hold even when there are serial correlations in the errors. For example, consider the simple linear regression model

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

where $u_t = e_t + \theta e_{t-1}$. Then,

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{t=1}^{n} (x_t - \bar{x}) e_t / n + \theta \sum_{t=1}^{n} (x_t - \bar{x}) e_{t-1} / n}{\sum_{t=1}^{n} (x_t - \bar{x})^2 / n}.$$

Consistency of OLS in simple linear regressions

- Thus, if
  1. $\sum_{t=1}^{n}(x_t - \bar{x})^2/n \xrightarrow{p} M \ (> 0)$
  2. $\bar{x} \xrightarrow{p} L \ (\in R)$
  3. $Ex_t^2 < c \ (\in R)$ for all t
  4. $$E\left(e_t | all\ x's\right) = 0 \text{ for all } t.$$

  and
  $$\begin{cases} Var\left(e_t | all\ x's\right) = \sigma^2 \text{ for all } t = 1, 2, \cdots, n \\ Cov\left(e_t,\ e_s | all\ x's\right) = 0 \text{ for all } s \neq t. \end{cases},$$

  $\hat{\beta}_1$ is consistent.

Large sample inference

- Recall that under the assumption of normality for the errors, the sampling distributions are normal, so we could derive t and F distributions for testing.
- This assumption of normal errors implied that the distribution of y, given the x's, was normal as well.
- It is easy to come up with examples for which this exact normality assumption will fail. Any clearly skewed variable, like wages, arrests, savings, etc. can't be normal, since a normal distribution is symmetric.

Large sample inference

Assume as before

1. $\{x_{it}\}$ is a sequence of random variables that satisfies some other conditions.
2. x_{it} is not linearly related to x_{jt} for any i and $j (\neq i)$. (No redundant information in regressors)
3. Zero conditional mean of the disturbance

$$E\left(u_t | all\ x's\right) = 0 \text{ for all } t.$$

Whatever values $x_{11}, x_{12}, ...x_{k(n-1)}, x_{kn}$ take, the mean of u_t is zero. This assumption implies

$$E\left(u_t\right) = 0 \text{ and } Cov\left(u_t, x_{jt}\right) = 0 \text{ for any } t \text{ and } j.$$

Large sample inference

4. Spherical disturbances

$$\begin{cases} Var\left(u_t \middle| all\ x's\right) = \sigma^2 \text{ for all } t = 1, 2, \cdots, n \\ Cov\left(u_t,\ u_s \middle| all\ x's\right) = 0 \text{ for all } s \neq t. \end{cases}$$

The assumption of common variance for u_t is called homoskedasticity.

- Under these assumptions, we have

$$\left(\hat{\beta}_j - \beta_j\right) / \sqrt{Var(\hat{\beta}_j)} \simeq N\left(0, 1\right)\right). \tag{1}$$

where $Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{t=1}^{n}(x_{jt} - \bar{x}_j)^2(1 - R_j^2)}$. That is, $\hat{\beta}$ is approximately normal with mean β and variance $\frac{\sigma^2}{\sum_{t=1}^{n}(x_{jt} - \bar{x}_j)^2(1 - R_j^2)}$.

- Here "$\simeq$" implies "convergence in distribution."

Large sample inference

- Let $F_n(x)$ and $F(x)$ be distribution functions of X_n and X, respectively. If $F_n(x) \to F(x)$ at every continuity point x of F, F_n is said to converge weakly to F, written $F_n \Rightarrow F$. In this case, $\{X_n\}$ is said to converge in distribution to X where X is a random variable with distribution function F, written $X_n \xrightarrow{d} X$.

Example

Let $\{X_i, i \geq 1\}$ be a sequence of i.i.d. r.v.s with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2 \neq 0$. Then

$$\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

- Relation (1) can be written as

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{Var(\hat{\beta}_j)}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

Large sample inference

1. $X_n \xrightarrow{d} X$ and $g(\cdot)$ is continuous

$$\Rightarrow g(X_n) \xrightarrow{d} g(X).$$

(This is called the continuous mapping theorem)

2. Suppose that $Y_n \xrightarrow{d} Y$ and $X_n \xrightarrow{p} c$ (a constant). Then

$$X_n + Y_n \xrightarrow{d} c + Y$$

$$X_n Y_n \xrightarrow{d} cY$$

$$\frac{Y_n}{X_n} \xrightarrow{d} \frac{Y}{c} \text{ when } c \neq 0.$$

Example

If $X_n \xrightarrow{d} N(0,1)$, $X_n^2 \xrightarrow{d} \chi^2(1)$.

Large sample inference

- The t-ratio for the null hypothesis

$$H_0 : \beta_j = \beta_j^0,$$

is defined as

$$\frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\widehat{Var(\hat{\beta}_j)}}},$$

where $\widehat{Var(\hat{\beta}_j)} = \frac{s^2}{\sum_{t=1}^{n}(x_{jt} - \bar{x}_j)^2(1 - R_j^2)}$. When $n \longrightarrow \infty$,

$$\frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\widehat{Var(\hat{\beta}_j)}}} \xrightarrow{d} N(0, 1).$$

Large sample inference

- We deduce from the distribution of the t-test

$$P\left(\hat{\beta}_j - c_{\alpha/2}SE(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c_{\alpha/2}SE(\hat{\beta}_j)\right) \simeq 1-\alpha$$

The approximate 100(1-α)% confidence interval for β_j is

$$\left[\hat{\beta}_j - c_{\alpha/2}SE(\hat{\beta}_j),\ \hat{\beta}_j + c_{\alpha/2}SE(\hat{\beta}_j)\right],$$

where $c_{\alpha/2}$ is taken from $N(0,1)$.

- Consider the null hypothesis

$$H_0 : \beta_1 = \ldots = \beta_k = 0.$$

The $F-$test for this null is defined as

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}.$$

As $n \to \infty$,

$$kF \xrightarrow{d} \chi^2(k).$$

Large sample inference

- Consider the null hypothesis

$$H_0 : \beta_{k-q+1} = ... = \beta_k = 0.$$

Estimate the restricted linear regression model

$$y_t = \beta_0 + \beta_1 x_{1t} + ... + \beta_{k-q} x_{k-q,t} + u_t$$

and let the sum of squared residuals RSS_r. The sum of sqaured residuals from regressing y on $x_1, ..., x_k$ is denoted as RSS_{ur} (unrestricted sum of squared residuals).The $F-$test for this null is defined as

$$F = \frac{(RSS_r - RSS_{ur})\,/\,q}{RSS_{ur}\,/\,(n-k-1)}.$$

As $n \to \infty$,

$$qF \xrightarrow{d} \chi^2(q).$$