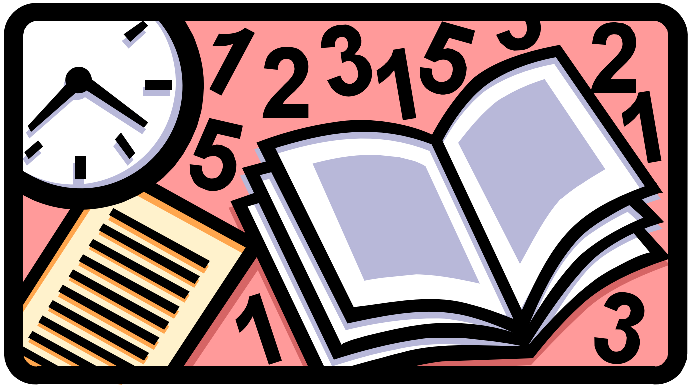
計量經濟分析(二)

政治大學統計系余清祥

2006年3月6日

第二週：大樣本理論、

OLS的估計問題



Email: csyue@nccu.edu.tw

大樣本理論

- 當樣本數較少時，資料蘊含的資訊不見得容易發掘；隨著樣本數的增加，資料背後的特性逐漸顯露。

→大樣本理論保證使用者，

在樣本數較大時的統計性質。

Q1. 你/妳覺得哪一個大樣本理論最重要？

Q2. 大樣本理論與一般不同之處？

(e.g.無限大是否大於無限大？)

假設與定義（一）

- 關於數列的定義：

  1.數列收斂(Convergence of sequence)

  2.有界(Bounded)數列

  3.大O (big o)及小O (little o)

  →舉例而言：

$$a_n = O(n^{\lambda}) \Leftrightarrow n^{-\lambda} a_n \text{ is bounded}$$

$$a_n = o(n^{\lambda}) \Leftrightarrow n^{-\lambda} a_n \to 0 \text{ as } n \to \infty$$

假設與定義（二）

- 隨機變數的定義

1. 機率收斂(Converge in probability)

→ $P(|X_n - a|) < \varepsilon \ as\ n \to \infty$, i.e., $X_n \xrightarrow{p} a$

$(Or,\ p\lim X_n = a)$

2. 機率有界(Bounded in probability)

→ $P(|X_n - b(\varepsilon)|) < \varepsilon \ \forall n \geq N(\varepsilon)$, i.e., $X_n = O_p(1)$

隨機變數的定義(續)

3. 分配收斂(Converge in distribution)

→ $F_n(x) \to F(x)$ $as\ n \to \infty\ \forall x \in R,\ i.e.,\ X_n \xrightarrow{d} X$

$$(e.g.,\ CLT: \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0,1))$$

4. Converge with prob. approaching one (w.p.a.1)

→ $P(\Omega_n) \to 1\ as\ n \to \infty,\ \{\Omega_n : n = 1,2,\ldots\} \subseteq F$

$where\ (\Omega, F, P): Probability\ space$

隨機變數的定義(續)

5.弱大數法則(Weak law of large numbers, WLLN)

→ $If \;\; E(|X_i|) < \infty, then \; \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu \; where \; \mu = E(X_i)$

6. Lindeberg-Levy

→ $If \;\; E(X_i^2) < \infty \,\&\, E(X_i) = 0, then \; \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{d} N(0, B)$

$where \; B = Var(X_i)$

重要的大樣本結果

- 變數收斂存在強弱關係：

→機率收斂可推得分配收斂。

(練習：證明中央極限定理！)

→機率、分配收斂的加減乘除之組合。

(練習：證明Lemma 3.2 及3.3)

- Slusky's Theorem

$$g : R^K \to R^J \ \&\ X_n \xrightarrow{p} c \Rightarrow g(X_n) \xrightarrow{p} g(c)$$

$$(i.e., p\lim g(X_n) = g(p\lim X_n)\ if\ \ g\ is\ continuous$$

重要的大樣本結果(續)

- Slusky's 定理的應用範例(變數矩陣)

$$If\ \ Z_n \xrightarrow{p} A \Rightarrow (1) Z_n^{-1}\ exists\ w.p.a.1$$

$$(2) Z_n^{-1} \xrightarrow{p} A^{-1}\ or\ p\lim Z_n^{-1} = A^{-1}$$

- 變數的連續函數

$$If\ \ X_n \xrightarrow{d} X\ \&\ g: R^K \rightarrow R^J \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

$$e.g.,\ If\ \ X_n \xrightarrow{d} N(0,V) \Rightarrow A' X_n \xrightarrow{d} N(0, A'VA)$$

$$X_n' V^{-1} X_n \xrightarrow{d} \chi^2$$

統計量的大樣本性質

- 通常對統計量的要求為不偏(unbiased)，再加上在所有不偏統計量中尋求變異數最小者。有時不偏統計量在有限樣本下不見得存在(e.g., 二項分配中參數倒數θ^{-1})，接近不偏估計量成為替代方案：

→一致估計量(Consistent estimator)

$$\hat{\theta}_n \xrightarrow{p} \theta, then\ \hat{\theta}_n\ is\ a\ consistent\ estimator$$

註：Asymptotically unbiased

統計量的大樣本性質(續)

- 一致估計量與樣本數

$$\hat{\theta}_n - \theta = O_p(n^{-1/2}) \Rightarrow \hat{\theta}_n \text{ } is \text{ } a \sqrt{n} - consistent \text{ } estimator$$

- Asymptotically normal

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,V) \Leftrightarrow \hat{\theta}_n \text{ } is \sqrt{n} - asymptotically \text{ } normal$$

註：效率(Efficiency)是兩個統計量的變異數比值。

統計量的大樣本性質(續)

- 估計量的大樣本等價($\sqrt{n}$ -equivalent)

$\rightarrow Var(\tilde{\theta}-\hat{\theta})\ is\ positive\ \ semidefinite\ \&\ \sqrt{n}(\tilde{\theta}-\hat{\theta})=o_p(1)$

- Asymptotically independent

$\rightarrow$If a estimator $\tilde{\theta}$ can be partitioned into two estimators $\hat{\theta}_1$ & $\hat{\theta}_2$ with

$$Var(\tilde{\theta})=V=\begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}$$

統計量的大樣本性質(續)

- Delta Method!!

→If $\hat{\theta}_n$ is $\sqrt{n}$-asymptotically normal with $Var(\hat{\theta}_n) = V$ positive definite and g is a continuous and differentiable function, then

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta)\right) \xrightarrow{d} N\left(0, \nabla g(\theta) V \nabla g(\theta)'\right)$$

單一方程式的線性模型

- 單一方程式(Single-equation Linear model)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + u$$

→常見的假設條件為

$$E(u) = 0, \;\; Cov(X_j, u) = 0, \; j = 1, 2, \ldots, K$$

較強的條件為 $E(u \mid X_1, \cdots, X_K) = E(u \mid X) = 0$

(證明！)。在較強的條件下可得出

$$E(Y \mid X_1, \ldots, X_K) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$$

單一方程式的線性模型(續)

- 單一方程式可用於建立結構模型(Structural model)，進一步探討因果關係；其中結構模型通常透過經濟模型或不正式的推論得出。(需考慮是否可估計！)
- 單一方程式的誤差項可能包括無法觀測的變數(Omitted variables)或量測誤差(Measurement error)，使得參數估計值不一致、或是甚至無法估計。

參數估計的可能問題

- 如果自變數與模型誤差相關，則此變數稱為內生(Endogenous)變數。內生變數的產生原因一般分為以下三種，或是這三種來源的組合：

  →Omitted variables

  →Measure error

  →Simultaneity

OLS的大樣本理論

- OLS的大樣本理論與以下三個假設有關：

$$OLS.1: E(X'u) = 0$$

$$OLS.2: rank\ E(X'X) = K \quad (i.e., full\ \ rank)$$

$$OLS.3: E(u^2X'X) = \sigma^2 E(X'X),\ where\ \sigma^2 = E(u^2)$$

→滿足這三項假設後，所有參數可被估計(Identified)、且變異數為定值(Homo-skedasticity)。

OLS的大樣本理論(續)

- 如果滿足OLS.1 及 OLS.2，則可證明參數估計值為一致估計量，其中

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i^{'} x_i / n \right)^{-1} \left(\sum_{i=1}^{n} x_i^{'} y_i / n \right)$$

→缺乏任一項上述假設，參數將無法估計。

註：上述參數估計量是由動差法(Method of moment)的類比原則導出，因為

$$\hat{\beta} = (X'X)^{-1} X'Y \Leftrightarrow \beta = [E(X'X)]^{-1} E(X'Y)$$

OLS的大樣本理論(續)

- 如果滿足OLS.1 ~ OLS.3，則可證明參數估計值服從大樣本常態分配：

$$\sqrt{n}(\hat{\beta}-\beta)\overset{a}{\sim}N(0,\sigma^2 A^{-1})$$

其中$A = E(X'X)$。

→如果少了OLS.3，則變異數將不齊一(Hetero-skedasticity)，有些人以WLS修正，但經常會有效率較差、或是估計值不一致的問題。

OLS的大樣本理論(續)

- 變異數將不齊一以穩健異質修正標準誤(Heteroskedasticity-robust standard error)較為合適，其中變異數以下式取代：

$$A\,Var(\hat{\beta}) = (X'X)^{-1}\left(\sum_{i=1}^{n}\hat{u}_i^2 X_i' X_i\right)(X'X)^{-1}$$

→上述的變異數可用於單一自變數的t檢定，但不能用於多個變數的檢定。

OLS的大樣本理論(續)

- 多個變數的檢定可利用Restricted model vs. Full model的方式，也就是在比較原始的受限模型之卡方值、與新加入變數降低的誤差之卡方值，兩者在原始模型的虛無假設下服從F分配。

參數估計的可能問題

- Omitted variables

→可能使得參數估計值產生偏誤(biased)，可透過Proxy variable解決，但須滿足某些不易確定的假設。

- Measure error

→分為自變數、因變數的量測誤差。

- Simultaneity

→解聯立方程式。

參數估計的可能問題(續)

- Omitted variables(遺漏變數)

→結構方程式少了重要解釋變數(e.g. *q*)

$$E(Y|X_1,X_2,\cdots,X_K,q)=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_KX_K+\gamma q$$

例如：工資(wage)問題中遺漏變數可能是個人的能力，與其中的「教育程度」有關。

→將結構方程式改寫為

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_KX_K+\gamma q+v$$

$$E(v|X_1,X_2\cdots,X_K,q)=0$$

- 註：上述的v為結構誤差，與之前誤差u滿足

$$u \equiv \gamma q + v$$

→如果遺漏變數與其他解釋有關聯，則這些解釋變數的OLS估計值可能為偏誤(OLS omitted variables inconsistency or OLS omitted variable bias)。

→估計值偏誤的解決方法之一為代理變數(Proxy variable)，可以消除或減少遺漏變數的影響。

- 代理變數Z需滿足兩個條件：

(1)代理變數必須是多餘(Redundant)的，即

$$E(Y \mid X, q, Z) = E(Y \mid X, q)$$

(2) 加入代理變數後即可去除遺漏變數的效果：

$$L(q \mid X, Z) = L(q \mid Z)$$

或是 $q = \theta_0 + \theta_1 Z + r$, 其中 $E(r) = 0, Cov(Z, r) = 0$ 。

→加入代理變數後經OLS求得為不偏估計量：

$$Y = \beta_0 \qquad\qquad + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \gamma q + v$$

$$Y = (\beta_0 + \gamma\theta_0) + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \gamma\theta_1 Z + (\gamma r + v)$$

- 如果加入了代理變數後的誤差 r 與解釋變數 X 有關，則稱為不完美的代理變數，其中 $q=\theta_0+\rho_1 X_1+\cdots+\rho_K X_K+\theta_1 Z+r$

→因此可推得OLS的估計值滿足

$$p\lim \hat{\beta}_j=\beta_j+\gamma\rho_j,$$

也就是OLS估計值為偏誤估計量。

(如果上式中ρ_j的數值比β_j小，則代理變數可以減小遺漏變數的問題。)