- 到目前为止，我们讨论了多元线性回归模型的估计与推断：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- 其中，我们主要关注OLS估计在给定任意样本情况下的性质，例如：

  无偏性，方差，正态性等

- 这些对任何样本量（n>k+1）都适用的性质叫做所谓有限样本性质、小样本性质或者精确性质。

- 这一章中，我们关心大样本（large sample）性质或者渐近（asymptotic）性质:样本量趋于无穷时OLS的性质

JEFFREY M. WOOLDRIDGE

Introductory Econometrics
A Modern Approach

SIXTH EDITION

Chapter 5

多元回归分析：OLS的渐近性

多元回归分析：OLS的渐近性

- **OLS有限样本性质**
  - 期望的无偏性：MLR.1 – MLR.4
  - 方差公式：MLR.1 – MLR.5
  - 高斯-马尔科夫定理：MLR.1 – MLR.5
  - 正态性：MLR.1 – MLR.6
- **本章将讨论的OLS大样本性质**
  - 一致性：MLR.1 – MLR.4
  - 渐近正态性：MLR.1 – MLR.5

无需假设误差项的正态性！

（思考：为什么需要大样本性质？）

一致性

- **一致性（consistency）**

估计量 θ_n 是总体参数 θ 的一致估计，如果随着 $n \to \infty$

$P(|\theta_n - \theta| < \epsilon) \to 1$ 对于任意 $\epsilon > 0$.

另一种符号：$plim\ \theta_n = \theta$

估计量依概率收敛于真实总体值

- **解释：**
  - 一致性意味着，通过增加样本量，可以使估计量任意接近（arbitrarily close）真实总体值的概率任意高（arbitrarily high）。
  - 一个充分条件： E(θ_n) 等于或者趋于θ ，Var(θ_n)趋于零

- **一致性是合理估计的最低要求**

一致性

- **定理 5.1 (OLS的一致性)**

$$MLR.1\text{–}MLR.4 \quad \Rightarrow \quad plim\ \hat{\beta}_j = \beta_j, \quad j = 0, 1, \ldots, k$$

- **简单回归模型的特例**

$$plim\ \hat{\beta}_1 = \beta_1 + \boxed{Cov(x_1, u)}/Var(x_1)$$

可以看出，如果解释变量是外生的，即与误差项不相关，则斜率估计是一致的。

- **假定 MLR.4‘**

$$E(u) = 0$$

$$Cov(x_j, u) = 0$$

所有解释变量必须与误差项不相关。该假设弱于零条件平均假设MLR.4.

2．随机序列的收敛

定义 随机序列$\{x_n\}_{n=1}^{\infty}=\{x_1,x_2,x_3,\cdots\}$“**依概率收敛**”(converges in probability)于常数 a，记为$\operatorname*{plim}_{n\to\infty} x_n=a$，或$x_n \xrightarrow{p} a$，如果$\forall \varepsilon>0$，当$n\to\infty$时，都有$\lim_{n\to\infty}\mathrm{P}\left(\left|x_n-a\right|>\varepsilon\right)=0$。

任意给定$\varepsilon>0$，当n越来越大时，随机变量x_n落在区间$(a-\varepsilon, a+\varepsilon)$之外的概率收敛于 0。

定义 随机序列$\{x_n\}_{n=1}^{\infty}$“**依概率收敛**”于随机变量x，记为$x_n \xrightarrow{p} x$，如果随机序列$\{x_n - x\}_{n=1}^{\infty}$依概率收敛于 0。

命题(连续函数与依概率收敛可交换运算次序，preservation of convergence for continuous transformation) 假设$g(\cdot)$为连续函数，则$\underset{n\to\infty}{\mathrm{p}\lim}\, g(x_n) = g\left(\underset{n\to\infty}{\mathrm{p}\lim}\, x_n\right)$。

当x_n的分布越来越集中于某x附近时，$g(x_n)$的分布自然也就越来越集中于$g(x)$附近。

概率收敛算子$\underset{n\to\infty}{\mathrm{p}\lim}$与连续函数$g(\cdot)$可交换运算次序。期望算子 E 无此性质，一般$\mathrm{E}(x^2) \neq [\mathrm{E}(x)]^2$。

3．依均方收敛

定义　随机序列$\{x_n\}_{n=1}^{\infty}$“**依均方收敛**”(converges in mean square)于常数 a，如果$\lim\limits_{n\to\infty} \mathrm{E}(x_n) = a$，$\lim\limits_{n\to\infty} \mathrm{Var}(x_n) = 0$。

命题　依均方收敛是依概率收敛的充分条件。

证明：使用切比雪夫不等式(参见附录)。

当x_n的均值越来越趋于 a，方差越来越小并趋于 0 时，就有$\operatorname*{plim}\limits_{n\to\infty} x_n = a$，即在极限处$x_n$退化(degenerate)为常数 a。

此命题是依均方收敛概念的主要用途。

4．依分布收敛

定义 记随机序列$\{x_n\}_{n=1}^{\infty}$与随机变量x的累积分布函数(cdf)分别为$F_n(\cdot)$与$F(\cdot)$。

如果对于任意实数c，都有$\lim\limits_{n\to\infty} F_n(c) = F(c)$，则称随机序列$\{x_n\}_{n=1}^{\infty}$“**依分布收敛**”(converge in distribution)于随机变量 x，记为$x_n \xrightarrow{d} x$。

【例】当 t 分布的自由度越来越大时，其累积分布函数收敛于标准正态的累积分布函数。

如果x为正态分布，而$x_n \xrightarrow{d} x$，则称$\{x_n\}_{n=1}^{\infty}$为“渐近正态”(asymptotically normal)。

依分布收敛意味着，两个随机变量的概率密度长得越来越像。

“依概率收敛”比“依分布收敛”更强(前者是后者的充分条件)：

$$\text{“}x_n \xrightarrow{p} x\text{”} \quad \Rightarrow \quad \text{“}x_n \xrightarrow{d} x\text{”}$$

反之不然：当x_n与x的分布函数很接近时，x_n与x的实际取值仍然可以很不相同(比如，x_n与x相互独立)。

命题 假设$g(\cdot)$为连续函数，且$x_n \xrightarrow{d} x$，则$g(x_n) \xrightarrow{d} g(x)$。

当x_n的分布越来越像x的分布时，$g(x_n)$的分布自然也越来越像$g(x)$的分布。

例： 假设$x_n \xrightarrow{d} z$，其中$z \sim N(0,1)$，

则$x_n^2 \xrightarrow{d} z^2$，其中$z^2 \sim \chi(1)$，即$x_n^2 \xrightarrow{d} \chi(1)$
(因为平方是连续函数)

渐近标准正态的平方服从渐近$\chi(1)$分布。

5.3 大数定律与中心极限定理

1．弱大数定律(Weak Law of Large Numbers)

假定$\{x_n\}_{n=1}^{\infty}$为独立同分布的随机序列，且$\mathrm{E}(x_1)=\mu$，$\mathrm{Var}(x_1)=\sigma^2$存在，则样本均值$\bar{x}_n \equiv \dfrac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{p} \mu$。

证明： 因为$\mathrm{E}(\bar{x}_n)=\mu$，而

$\mathrm{Var}(\bar{x}_n)=\mathrm{Var}\left(\dfrac{x_1+\cdots+x_n}{n}\right)=\dfrac{1}{n^2}n\sigma^2=\dfrac{\sigma^2}{n}\to 0$，故$\bar{x}_n$依均方收敛于$\mu$。因此，$\bar{x}_n \xrightarrow{p} \mu$。样本无限大时，样本均值趋于总体均值，故名“大数定律”。

2．中心极限定理(Central Limit Theorem，简记 CLT)

定理　假定$\{x_n\}_{n=1}^{\infty}$为独立同分布的随机序列，且$\mathrm{E}(x_1)=\mu$，$\mathrm{Var}(x_1)=\sigma^2$存在，则$\sqrt{n}\,(\bar{x}_n-\mu)\xrightarrow{d}N(0,\sigma^2)$。

根据弱大数定律，$(\bar{x}_n-\mu)\xrightarrow{p}0$，而$\sqrt{n}\to\infty$，故用$\sqrt{n}\,(\bar{x}_n-\mu)$(即"$\infty\cdot 0$"型)得到非退化分布。

进一步，$(\bar{x}_n-\mu)$收敛到 0 的速度与$\dfrac{1}{\sqrt{n}}$收敛到 0 的速度类似(二者乘积为非退化分布)，称为"$\sqrt{n}$收敛"(root-n convergence)。

直观上，可视为$\bar{x}_n\xrightarrow{d}N\left(\mu,\dfrac{\sigma^2}{n}\right)$；但不严格，因为$\dfrac{\sigma^2}{n}\to 0$。

在一维情况下，中心极限定理可等价地写为$\dfrac{\overline{x}_n-\mu}{\sqrt{\sigma^2/n}}\xrightarrow{d}N(0,1)$，但此形式不易推广到多维的情形。

推广到多维的情形：

假定$\{\boldsymbol{x}_n\}_{n=1}^{\infty}$为独立同分布的随机向量序列，且$\mathrm{E}(\boldsymbol{x}_1)=\boldsymbol{\mu}$，$\mathrm{Var}(\boldsymbol{x}_1)=\boldsymbol{\Sigma}$存在，则$\sqrt{n}(\overline{\boldsymbol{x}}_n-\boldsymbol{\mu})\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Sigma})$。

2．一致估计量

定义　如果$\underset{n\to\infty}{\mathrm{plim}}\,\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}$，则估计量$\hat{\boldsymbol{\beta}}_n$是参数$\boldsymbol{\beta}$的**一致估计量**(consistent estimator)。

一致性(consistency)意味着，当样本容量足够大时，$\hat{\boldsymbol{\beta}}_n$依概率收敛到真实参数$\boldsymbol{\beta}$。

这是对估计量最基本，也是最重要的要求。

如果估计方法不一致，意味着研究没有太大意义；因为无论样本容量多大，估计量也不会收敛到真实值。

3．渐近正态分布与渐近方差

定义　如果$\sqrt{n}\,(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$，其中$\boldsymbol{\Sigma}$为半正定矩阵，则称$\hat{\boldsymbol{\beta}}_n$为**渐近正态分布**(asymptotically normally distributed)，称$\boldsymbol{\Sigma}$为**渐近方差**(asymptotic variance)，记为$\mathrm{Avar}(\hat{\boldsymbol{\beta}}_n)$。

可近似地认为$\hat{\boldsymbol{\beta}}_n \xrightarrow{d} N\left(\boldsymbol{\beta}, \boldsymbol{\Sigma}/n\right)$。$(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$收敛到 0 的速度与$\dfrac{1}{\sqrt{n}}$收敛到 0 的速度相同，称为“$\sqrt{n}$收敛”(root-n convergence)。

4．渐近有效

假设$\hat{\boldsymbol{\beta}}_n$与$\tilde{\boldsymbol{\beta}}_n$都是$\boldsymbol{\beta}$的渐近正态估计量，其渐近方差分别为$\boldsymbol{\Sigma}$与$\boldsymbol{V}$。如果$(\boldsymbol{V} - \boldsymbol{\Sigma})$为半正定矩阵，则称$\hat{\boldsymbol{\beta}}_n$比$\tilde{\boldsymbol{\beta}}_n$**更为渐近有效**(asymptotically more efficient)。

5.5 渐近分布的推导

推导渐近分布的常用技巧，涉及依概率收敛与依分布收敛的交叉运算，统称“斯拉斯基定理”(Slutsky Theorem)。

(1) $x_n \xrightarrow{d} x,\ y_n \xrightarrow{p} a \ \Rightarrow\ x_n + y_n \xrightarrow{d} x + a$。

在极限处，y_n退化为常数a，故$x_n + y_n$在极限处只是将x_n的渐近分布x位移到$x + a$。

特例：如果$a = 0$，则$x_n + y_n \xrightarrow{d} x$。

(2) $x_n \xrightarrow{d} x,\ y_n \xrightarrow{p} 0 \ \Rightarrow\ x_n y_n \xrightarrow{p} 0$。

在极限处，y_n退化为 0，x_n有正常的渐近分布x，故$x_n y_n$退化为 0。

(3) 随机向量$\boldsymbol{x}_n \xrightarrow{d} \boldsymbol{x}$，随机矩阵$\boldsymbol{A}_n \xrightarrow{p} \boldsymbol{A}$，$\boldsymbol{A}_n \boldsymbol{x}_n$可以相乘 $\Rightarrow \boldsymbol{A}_n \boldsymbol{x}_n \xrightarrow{d} \boldsymbol{A}\boldsymbol{x}$。

特例：如果$\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$，则$\boldsymbol{A}_n \boldsymbol{x}_n \xrightarrow{d} N(\mathbf{0}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$。

在极限处，随机矩阵$\boldsymbol{A}_n$退化为常数矩阵$\boldsymbol{A}$。

正态分布的线性组合仍服从正态分布，且$\mathrm{Var}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}\,\mathrm{Var}(\boldsymbol{x})\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$。

(4) 随机向量$\boldsymbol{x}_n \xrightarrow{d} \boldsymbol{x}$,随机矩阵$\boldsymbol{A}_n \xrightarrow{p} \boldsymbol{A}$,$\boldsymbol{A}_n \boldsymbol{x}_n$可以相乘,$\boldsymbol{A}^{-1}$存在 $\Rightarrow$ 二次型$\boldsymbol{x}_n' \boldsymbol{A}_n^{-1} \boldsymbol{x}_n \xrightarrow{d} \boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x}$。

一致性

- 对于OLS的一致性，只需要较弱的MLR.4‘
- MLR.4‘不成立则很有可能导致OLS的一致性不成立
- 遗漏变量的渐近偏误

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v$$

真实模型

$$y = \beta_0 + \beta_1 x_1 + [\beta_2 x_2 + v] = \beta_0 + \beta_1 x_1 + u$$

误设模型

$$\Rightarrow \quad plim\ \tilde{\beta}_1 = \beta_1 + Cov(x_1, u)/Var(x_1)$$

偏误

$$= \beta_1 + \boxed{\beta_2 Cov(x_1, x_2)}/Var(x_1) = \beta_1 + \boxed{\beta_2 \delta_1}$$

如果遗漏变量与已含变量不相关，则没有遗漏变量偏误

渐近正态和大样本推断

- **渐近正态和大样本推断**
  - 在现实中，正态性假定MLR.6常常有问题
  - 如果MLR.6不能成立，则t或F统计量可能是错的
  - 幸运的是，即使没有MLR.6，OLS估计在大样本下也是正态分布的
  - 这意味着，如果样本足够大，t或F检验仍然有效

- **定理5.2 （OLS的渐近正态性** Asymptotic normality）

在假定MLR.1 – MLR.5下:

$$\frac{(\hat{\beta}_j - \beta_j)}{se(\hat{\beta}_j)} \overset{a}{\sim} \text{Normal}(0, 1)$$

在大样本中，标准化的
估计量是正态分布的

而且 $plim\ \hat{\sigma}^2 = \sigma^2$

渐近正态和大样本推断

- **注意：**
  - MLR.1-MLR.5仍然是需要的，例如，异方差性
  - $\hat{\beta}_j-\beta_j$**不是**渐近正态！根据一致性，$\hat{\beta}_j-\beta_j$与$se(\hat{\beta}_j)$均趋于零。
  - 可以证明$\sqrt{\mathrm{n}}(\hat{\beta}_j-\beta_j)$是渐近正态分布
  - 与有限样本正态性对比：

    大样本：在假定MLR.1 – MLR.5下：

$$\frac{(\hat{\beta}_j-\beta_j)}{se(\hat{\beta}_j)} \overset{a}{\sim} \text{Normal}(0,1)$$

有限样本：在假定MLR.1 – MLR.6下：

$$\frac{\hat{\beta}_j-\beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

注意：是否矛盾?

渐近正态和大样本推断

- **实际分析中**
  - 在大样本下，t分布接近于标准正态分布
  - 即使没有MLR.6，t检验在大样本下也是有效的
  - 置信区间与F检验也是有效的

- **OLS抽样误差的渐近性分析**

$$\widehat{Var}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{SST_j(1-R_j^2)}$$

收敛到 σ^2

收敛到 $n \cdot Var(x_j)$

收敛到一个固定值

渐近正态和大样本推断

- **OLS抽样误差的渐近性分析（续）**

$\widehat{Var}(\hat{\beta}_j)$ 以 $1/n$ 的速度收缩

$se(\hat{\beta}_j)$ 以 $\sqrt{1/n}$ 的速度收缩

- **这就是为什么大样本更好的原因**
- **例子：出生体重方程中的标准误**

$$n = 1,388 \ \Rightarrow se(\hat{\beta}_{cigs}) = .00086$$

$$n = 694 \ \Rightarrow se(\hat{\beta}_{cigs}) = .0013$$

$$\frac{.00086}{.0013} \approx \sqrt{\frac{694}{1,388}}$$

只使用观测值的前半部分