- 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 $n \to \infty$

$$P\left(|\theta_n - \theta| < \epsilon\right) \to 1 \text{ 对于任意 } \epsilon > 0.$$

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)**

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

- **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.

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

误设模型

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

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

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**

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

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

只使用观测值的前半部分

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