第 5 章　大样本 OLS

5.1　为何需要大样本理论

“大样本理论”(large sample theory)，也称“渐近理论”(asymptotic theory)，研究当样本容量 n 趋无穷时统计量的性质。

大样本理论近年来大受欢迎的原因如下。

(1) 小样本理论的假设过强。小样本理论的严格外生性假设要求解释变量与所有的扰动项均正交。在时间序列模型中，这意味着解释变量与扰动项的过去、现在与未来值全部正交！

自回归模型必然违背此假定。大样本理论只要求解释变量与同期扰动项不相关。

例　$y_t = \beta y_{t-1} + \varepsilon_t$，其中$\mathrm{E}(y_{t-1}\varepsilon_t) = 0$。

由于ε_t是y_t的一部分，故二者相关，即

$$\mathrm{E}(y_t\varepsilon_t) = \mathrm{E}\left[(\beta y_{t-1} + \varepsilon_t)\varepsilon_t\right] = \beta\,\mathrm{E}(y_{t-1}\varepsilon_t) + \mathrm{E}(\varepsilon_t^2) = \mathrm{E}(\varepsilon_t^2) > 0$$

小样本理论假定扰动项为正态分布，大样本理论无此限制。

(2) 小样本的精确分布(exact distribution)难推导。大样本的渐近分布较易推导。

(3) 大样本理论要求样本容量较大，至少$n \geq 30$，最好 100 以上。

5.2 随机收敛

1．确定性序列的收敛

定义 确定性序列$\{a_n\}_{n=1}^{\infty}=\{a_1,a_2,a_3,\cdots\}$ “**收敛**” (converges)于常数 a，记为$\lim\limits_{n\to\infty}a_n=a$或$a_n\to a$，如果$\forall\varepsilon>0$，存在$N>0$，只要$n>N$，就有$|a_n-a|<\varepsilon$，即$\{a_{N+1},a_{N+2},\cdots\}$均落入区间$(a-\varepsilon,a+\varepsilon)$内。

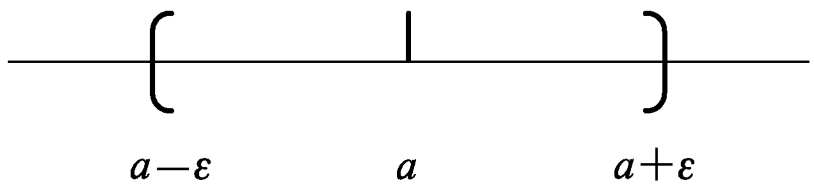

图 5.1 确定性序列的收敛

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$\lim_{n\to\infty}\mathrm{P}\left(|x_n-a|>\varepsilon\right)=0$。

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。

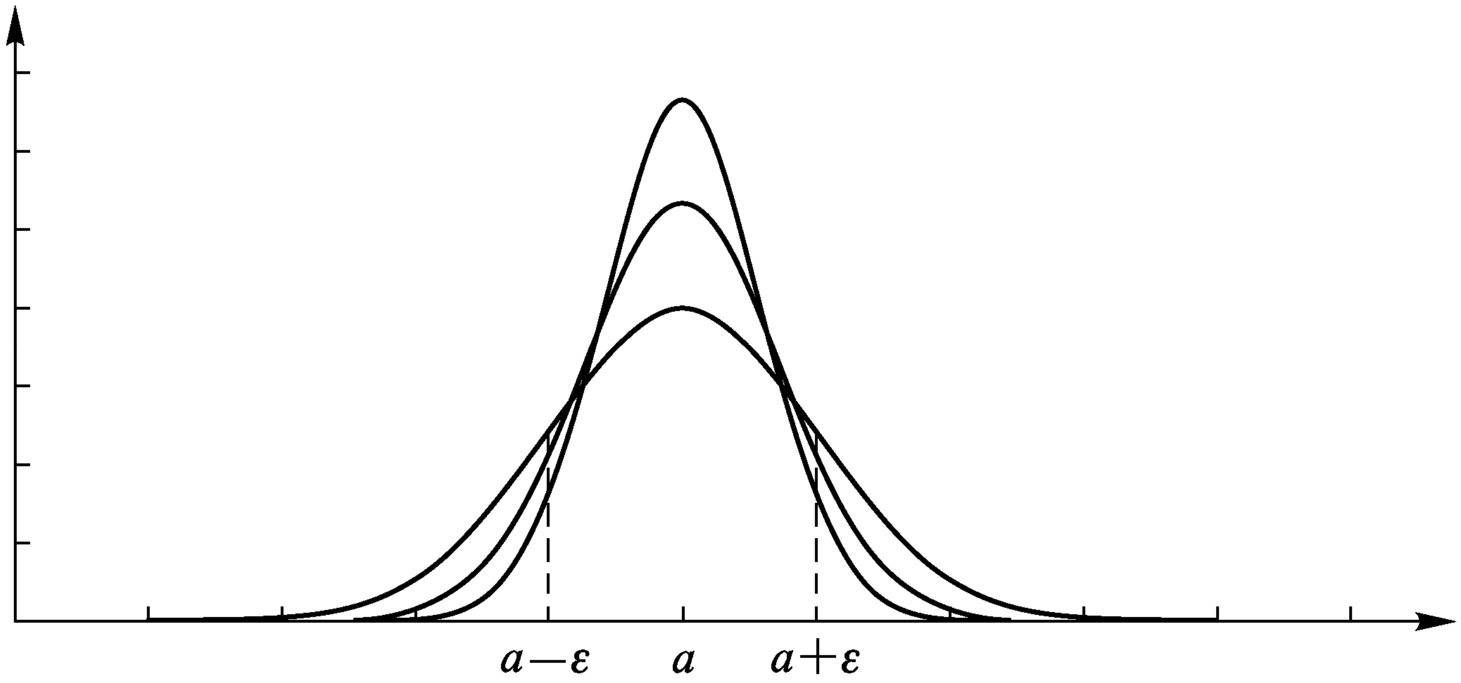

图 5.2　随机序列的收敛

对于随机向量与随机矩阵，也可定义依概率收敛，只要定义其每个元素都依概率收敛即可。

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$\operatorname*{p\,lim}_{n\to\infty} g(x_n) = g\left(\operatorname*{p\,lim}_{n\to\infty} x_n\right)$。

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$\operatorname*{p\,lim}_{n\to\infty}$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$\mathrm{E}(x^2) \neq \left[\mathrm{E}(x)\right]^2$。

例： 如果$\operatorname*{plim}_{n\to\infty} s^2 = \sigma^2$，则

$$\operatorname*{plim}_{n\to\infty} s = \operatorname*{plim}_{n\to\infty}(s^2)^{1/2} = (\operatorname*{plim}_{n\to\infty} s^2)^{1/2} = (\sigma^2)^{1/2} = \sigma$$

(因为开根号是连续函数）。

如果样本方差是方差的一致估计，则样本标准差也是标准差的一致估计)。

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$\lim_{n\to\infty} F_n(c) = F(c)$，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。

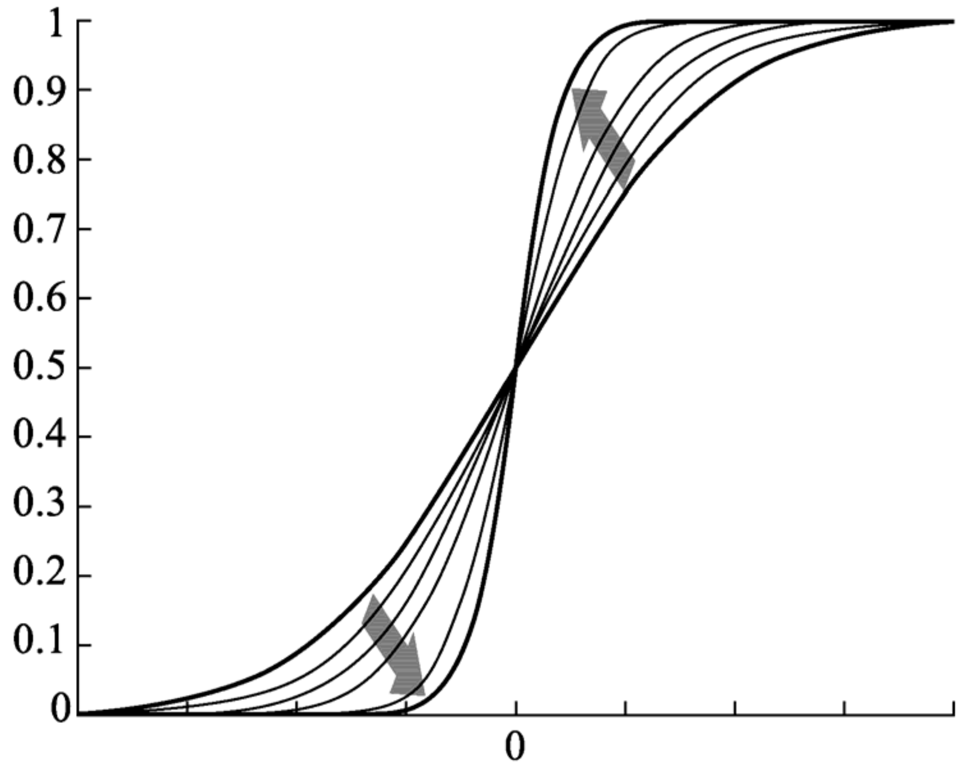

图 5.3　依分布收敛

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$\bar{x}_n \equiv \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{p} \mu$。

证明： 因为$\mathrm{E}(\bar{x}_n)=\mu$，而

$\mathrm{Var}(\bar{x}_n)=\mathrm{Var}\left(\frac{x_1+\cdots+x_n}{n}\right)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}\to 0$，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$\sqrt{n}\,(\overline{x}_n-\mu)\xrightarrow{d}N(0,\sigma^2)$。

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

进一步，$(\overline{x}_n-\mu)$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$\overline{x}_n\xrightarrow{d}N\left(\mu,\dfrac{\sigma^2}{n}\right)$；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$\dfrac{\bar{x}_n-\mu}{\sqrt{\sigma^2/n}}\xrightarrow{d}N(0,1)$，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$\sqrt{n}(\bar{\boldsymbol{x}}_n-\boldsymbol{\mu})\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Sigma})$。

5.4 统计量的大样本性质

1．均方误差

假设$\hat{\beta}$是一维参数β的估计量。希望抽样误差$(\hat{\beta}-\beta)$尽量地小。

定义 以估计量$\hat{\beta}$来估计参数β，则其**均方误差**(Mean Squared Error，简记 MSE)为

$$\mathrm{MSE}(\hat{\beta}) \equiv \mathrm{E}\left[(\hat{\beta}-\beta)^2\right]$$

一个最优的估计量应在所有估计量中均方误差最小。

不希望$\hat{\beta}$系统地高估或低估β，即无系统误差(systematic error)。

定义 以估计量$\hat{\beta}$来估计参数β，则其**偏差**为$\text{Bias}(\hat{\beta}) \equiv \text{E}(\hat{\beta}) - \beta$。

定义 如果偏差$\text{Bias}(\hat{\beta}) = 0$，则称$\hat{\beta}$为**无偏估计量**(unbiased estimator)。

命题 均方误差可分解为方差与偏差平方之和，即

$$\text{MSE}(\hat{\beta}) = \text{Var}(\hat{\beta}) + \left[\text{Bias}(\hat{\beta})\right]^2$$

证明： $\text{MSE}(\hat{\beta}) \equiv \text{E}\left[(\hat{\beta} - \beta)^2\right] = \text{E}\left\{\left[\hat{\beta} - \text{E}(\hat{\beta}) + \text{E}(\hat{\beta}) - \beta\right]^2\right\}$

$$= \mathrm{E}\left[\hat{\beta}-\mathrm{E}(\hat{\beta})\right]^2 + 2\mathrm{E}\left\{\left[\hat{\beta}-\mathrm{E}(\hat{\beta})\right]\left[\mathrm{E}(\hat{\beta})-\beta\right]\right\} + \mathrm{E}\left[\mathrm{E}(\hat{\beta})-\beta\right]^2$$

$$= \mathrm{Var}(\hat{\beta}) + 2\mathrm{E}\left\{\left[\hat{\beta}-\mathrm{E}(\hat{\beta})\right]\left[\mathrm{E}(\hat{\beta})-\beta\right]\right\} + \left[\mathrm{Bias}(\hat{\beta})\right]^2$$

上式的交叉项为

$$\mathrm{E}\left\{\left[\hat{\beta}-\mathrm{E}(\hat{\beta})\right]\left[\mathrm{E}(\hat{\beta})-\beta\right]\right\} = \left[\mathrm{E}(\hat{\beta})-\beta\right]\mathrm{E}\left[\hat{\beta}-\mathrm{E}(\hat{\beta})\right] = \left[\mathrm{E}(\hat{\beta})-\beta\right]\cdot 0 = 0$$

均方误差最小化，可视为在“估计量方差”与“偏差”之间进行权衡(trade-off)。

多维情形的类似结论：

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}) \equiv \mathrm{E}\left[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\right] = \mathrm{Var}(\hat{\boldsymbol{\beta}}) + \left[\mathrm{Bias}(\hat{\boldsymbol{\beta}})\right]\left[\mathrm{Bias}(\hat{\boldsymbol{\beta}})\right]'$$

2．一致估计量

定义　如果 $\operatorname*{plim}_{n\to\infty}\hat{\boldsymbol{\beta}}_n=\boldsymbol{\beta}$，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$\text{Avar}(\hat{\boldsymbol{\beta}}_n)$。

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 $\text{Var}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}\,\text{Var}(\boldsymbol{x})\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$。

(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$\boldsymbol{x}_n' \boldsymbol{A}_n^{-1} \boldsymbol{x}_n \xrightarrow{d} \boldsymbol{x}' \boldsymbol{A}^{-1} \boldsymbol{x}$。

5.6　随机过程的性质

随机序列$\{x_n\}_{n=1}^{\infty}$也称“随机过程”(stochastic process)。如下标为时间，记为$\{x_t\}_{t=1}^{\infty}$，也称“时间序列”(time series)。

1．严格平稳过程

考察中国 1978—2007 年的通货膨胀率，即$\{\pi_{1978}, \pi_{1979}, \cdots, \pi_{2007}\}$，假如每年的通货膨胀率作为随机变量都有不同的分布，如何估计$\mathrm{E}(\pi_{1978})$与$\mathrm{Var}(\pi_{1978})$？每年通货膨胀率的样本容量仅为 1！

如果 30 年的通货膨胀率分布都不变，可将$\bar{\pi} \equiv \dfrac{1}{30}\sum_{t=1978}^{2007} \pi_t$作为$\mathrm{E}(\pi_t)$的估计量。

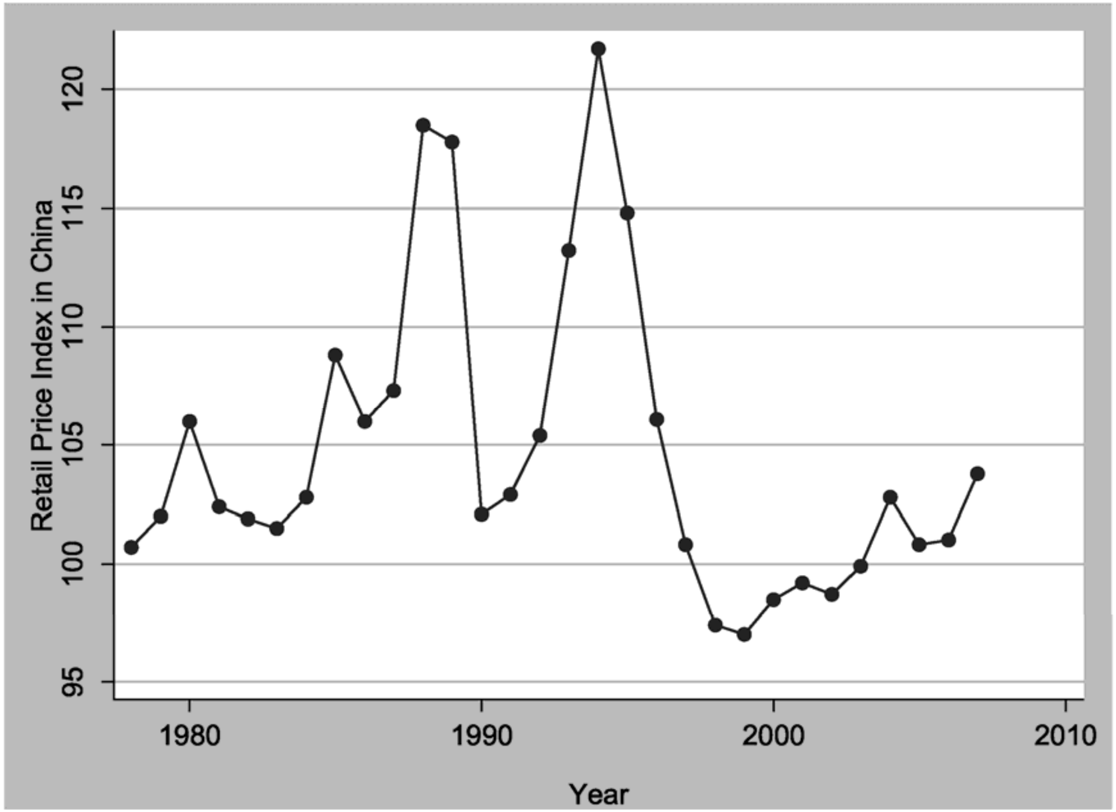

图 5.4　中国零售物价环比指数，1978—2007

“严格平稳过程”要求有限维分布不随时间推移而改变。

【例】x_i的分布与x_j的分布相同$(\forall i, j)$；

(x_1, x_4)的分布与(x_2, x_5)相同；

(x_1, x_2, x_3)的分布与(x_5, x_6, x_7)相同。

定义 随机过程$\{x_t\}_{t=1}^{\infty}$是**严格平稳过程**(strictly stationary process)，简称平稳过程，如果对任意 m 个时期的时间集合$\{t_1, t_2, \cdots, t_m\}$，随机向量$\{x_{t_1}, x_{t_2}, \cdots, x_{t_m}\}$的联合分布等于随机向量$\{x_{t_1+k}, x_{t_2+k}, \cdots, x_{t_m+k}\}$的联合分布，其中$k$为任意整数。

$\{x_{t_1}, x_{t_2}, \cdots, x_{t_m}\}$的联合分布仅依赖于$\{t_1, t_2, \cdots, t_m\}$各个时期之间的相对距离，而不依赖于其绝对位置。

例　如果随机过程$\{x_t\}_{t=1}^{\infty}$为 iid，则$\{x_t\}_{t=1}^{\infty}$是平稳过程，且不存在序列相关。

例　如果随机过程$\{x_t\}_{t=1}^{\infty} = \{x_1, x_1, x_1, \cdots\}$(即$x_t \equiv x_1$)，则$\{x_t\}_{t=1}^{\infty}$是平稳过程，且存在最强的序列相关。

例　考虑以下一阶自回归过程(first order autoregression，简记AR(1))，

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad \mathrm{Cov}(y_{t-1}, \varepsilon_t) = 0$$

其中，$\{\varepsilon_t\}$为独立同分布。

命题 如果$\rho=1$，则$\{y_t\}$不是平稳过程。如果$|\rho|<1$，则$\{y_t\}$是严格平稳过程。

证明：如果$\rho=1$，则$y_t = y_{t-1}+\varepsilon_t$。因此，$y_t = y_0+\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_t$。

故当$t\to\infty$时，$\mathrm{Var}(y_t)=t\sigma_\varepsilon^2\to\infty$，其中$\sigma_\varepsilon^2\equiv\mathrm{Var}(\varepsilon_t)$，即方差越来越大，以至无穷。因此，$\{y_t\}$不是平稳过程。此时，$\{y_t\}$被称为“随机游走”(random walk)，存在“单位根”(unit root)。

如果$|\rho|<1$，对该方程两边同时取方差，可得

$$\mathrm{Var}(y_t)=\rho^2\,\mathrm{Var}(y_{t-1})+\sigma_\varepsilon^2$$

这是一阶线性差分方程。由于$\rho^2<1$，故$\mathrm{Var}(y_t)$将收敛于$\dfrac{\sigma_\varepsilon^2}{1-\rho^2}$。

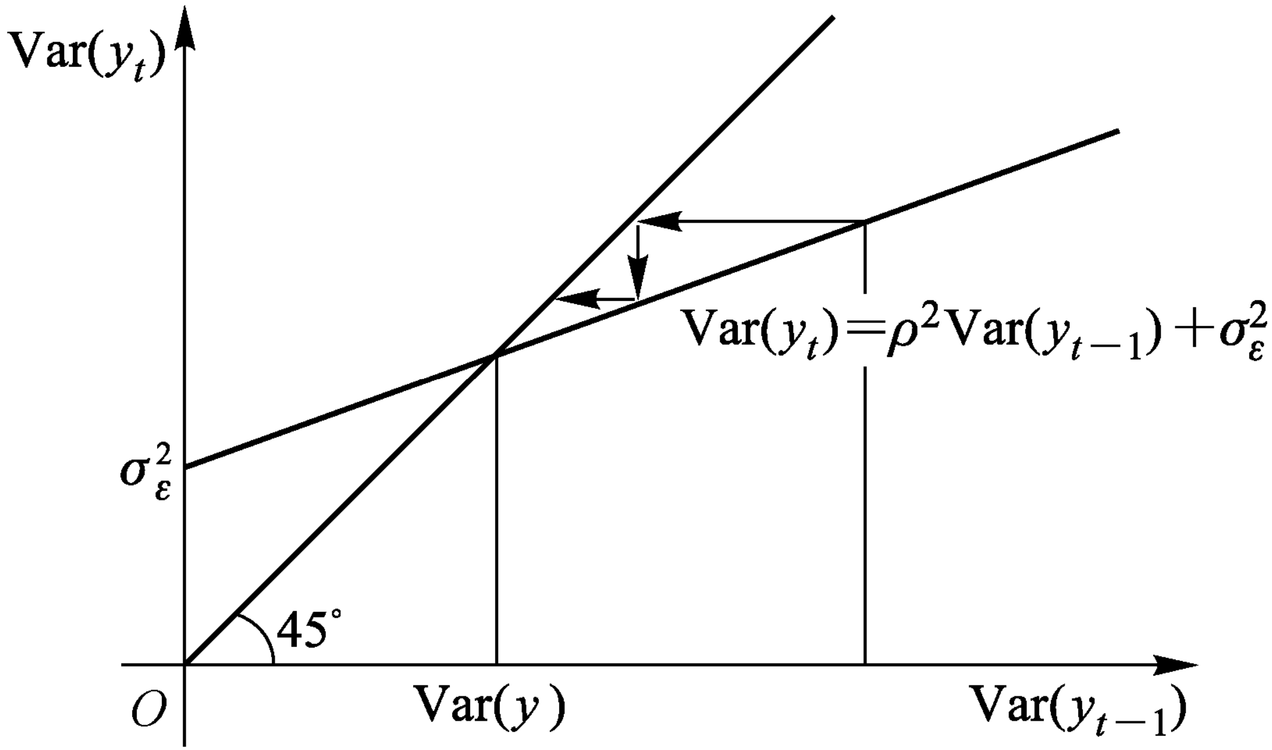

图 5.5　平稳一阶自回归过程的方差收敛

定义　随机过程$\{x_t\}_{t=1}^{\infty}$是**弱平稳过程**(weakly stationary process)或**协方差平稳过程**(covariance stationary process)，如果$\mathrm{E}(x_t)$不依赖于t，而且$\mathrm{Cov}(x_t, x_{t+k})$仅依赖于$k$(即$x_t$与$x_{t+k}$在时间上的相对距离)而不依赖于其绝对位置$t$。

弱平稳过程的期望与方差均为常数。

在$\mathrm{Cov}(x_t, x_{t+k})$中令$k=0$，可知方差为常数。

定义　一个协方差平稳过程$\{x_t\}_{t=1}^{\infty}$被称为**白噪声过程**(white noise process)，如果对于$\forall\, t$，都有$\mathrm{E}(x_t)=0$，而且$\mathrm{Cov}(x_t, x_{t+k})=0$，$\forall\, k\neq 0$。

注：白噪声过程不一定是 iid，也不一定严格平稳。

严格平稳过程是弱平稳过程的充分条件。

但反之则不然，因为弱平稳过程只要求二阶矩平稳(即期望、方差、协方差等不随时间而变)，而概率分布可能依赖于更高阶矩。

对于随机向量过程$\{\boldsymbol{x}_t\}_{t=1}^{\infty}$，可类似定义平稳过程或弱平稳过程。

如果$\{\boldsymbol{x}_t\}_{t=1}^{\infty}$为(弱)平稳过程，则其每个分量都是(弱)平稳过程；反之，则不然。

2．渐近独立性

“严格平稳过程”(相当于“同分布”假定)还不足以应用大数定律或中心极限定理，因为它们都要求独立同分布。

但相互独立的假定对于大多数经济变量而言过强。比如，今年的通胀率显然与去年的通胀率相关。

但今年的通胀率与 100 年前的通胀率或许可近似地视为相互独立，称为**渐近独立**(ergodic，也称“遍历性”)。

渐近独立意味着，只要两个随机变量相距足够远，可近似认为它们相互独立。

Ergodicity

- Ergodicity refers to the persistence of the sequence.
- Formally, we say that a stationary process $\{z_i\}$ is **ergodic** if, for any two bounded functions $f : \mathbb{R}^{k+1} \to \mathbb{R}$ and $g : \mathbb{R}^{l+1} \to \mathbb{R}$,

$$\lim_{n\to\infty} |\mathrm{E}[f(z_i, \ldots, z_{i+k}) \cdot g(z_{i+n}, \ldots, z_{i+n+l})]|$$
$$= |\mathrm{E}[f(z_i, \ldots, z_{i+k})]| \cdot |\mathrm{E}[g(z_{i+n}, \ldots, z_{i+n+l})]|$$

- We can understand this formula by considering a (slightly stronger) condition, for two scalar random variables, x and y.
- Since $\mathrm{Cov}(x, y) = \mathrm{E}[xy] - \mathrm{E}[x]\mathrm{E}[y]$, it follows that

$$\mathrm{E}[xy] = \mathrm{E}[x]\mathrm{E}[y] \iff \mathrm{Cov}(x, y) = 0$$

- If we further require that, for **any** functions f and g,

$$\mathrm{E}[f(x)g(y)] = \mathrm{E}[f(x)]\mathrm{E}[g(y)]$$

  then in effect, not only can't x and y be correlated, but they can't have any relationship whatsoever (and so they must be independent).
- Ergodicity means that this condition must be met in the limit as the random variables become far apart in the sequence.
- Intuitively, what this means is that a sequence is ergodic if it is asymptotically independent (two random variables positioned far apart in the sequence are almost independently distributed.)
- Stationary processes that are ergodic are called **ergodic stationary**.

例 AR(1)是否渐近独立？

考虑$y_t = \rho y_{t-1} + \varepsilon_t$，其中$|\rho| < 1$。当时间间隔为 1 时，

$$\mathrm{Cov}(y_t, y_{t-1}) = \mathrm{Cov}(\rho y_{t-1} + \varepsilon_t, y_{t-1}) = \rho\sigma_y^2$$

当时间间隔为 2 时，

$$y_t = \rho y_{t-1} + \varepsilon_t = \rho(\rho y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \rho^2 y_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t$$

故

$$\mathrm{Cov}(y_t, y_{t-2}) = \mathrm{Cov}(\rho^2 y_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t,\ y_{t-2}) = \rho^2\sigma_y^2$$

当时间间隔为j时，$\mathrm{Cov}(y_t, y_{t-j}) = \rho^j\sigma_y^2$。由于$|\rho| < 1$，故当$j \to \infty$时，

$\mathrm{Cov}(y_t, y_{t-j}) \to 0$。因此，AR(1)为渐近独立。

渐近独立定理(Ergodic Theorem)　假设$\{x_i\}_{i=1}^{\infty}$为渐近独立的严格平稳过程，且$\mathrm{E}(x_i) = \mu$，则$\bar{x}_n \equiv \dfrac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{p} \mu$，即样本均值$\bar{x}_n$是总体均值$\mathrm{E}(x_i)$的一致估计。

这是对大数定律的重要推广，更适用于经济数据。

大数定律要求每个x_i相互独立，而渐近独立定理允许$\{x_i\}_{i=1}^{\infty}$存在"序列相关"(serial correlation)，只要此相关关系在极限处消失。

大数定律要求每个x_i的分布相同，而渐近独立定理要求$\{x_i\}_{i=1}^{\infty}$为严格平稳过程(故也同分布)。

命题　如果$\{x_i\}_{i=1}^{\infty}$为渐近独立的严格平稳过程，则对于任何连续函数$f(\cdot)$，$\{f(x_i)\}_{i=1}^{\infty}$也是渐近独立的严格平稳过程。

渐近独立定理意味着，渐近独立平稳过程$\{x_i\}_{i=1}^{\infty}$的任何“总体矩”(population moment)$\mathrm{E}\left[f(x_i)\right]$，都可以由其对应的“样本矩”(sample moment)$\frac{1}{n}\sum_{i=1}^{n} f(x_i)$一致地估计。

【例】$\mathrm{E}\left(\boldsymbol{x}_i\boldsymbol{x}_i'\right)_{K\times K}$可由$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'$一致地估计，其中$\left(\boldsymbol{x}_i\boldsymbol{x}_i'\right)_{K\times K}$为随机矩阵。其中，$\boldsymbol{x}_i \equiv (x_{i1}\ x_{i2}\ \cdots\ x_{iK})'$。

使用中心极限定理还需另一条件，即鞅差分序列。

定义　称随机过程$\{x_i\}_{i=1}^{\infty}$为**鞅**(martingale)，如果它满足$\mathrm{E}(x_i \mid x_{i-1},\cdots,x_1)=x_{i-1}$，$\forall i \geq 2$。

例　随机游走过程$x_t = x_{t-1}+\varepsilon_t$。显然，$\mathrm{E}(x_t \mid x_{t-1},\cdots,x_1)=x_{t-1}$。

例　资本市场有效理论认为，所有有关未来价格的已知信息均已反映在当期价格上，故$\mathrm{E}(p_{t+1} \mid p_t,\cdots,p_1)=p_t$。

定义　称随机过程$\{x_i\}_{i=1}^{\infty}$为**鞅差分序列**(Martingale Difference Sequence，简记 MDS)，如果它满足$\mathrm{E}(x_i \mid x_{i-1}, \cdots, x_1) = 0$，$\forall i \geq 2$。

这意味着x_i均值独立于它的所有过去值。

因此，$\mathrm{Cov}(x_i, x_{i-j}) = 0$，$\forall j \neq 0$。

根据迭代期望定律可知，鞅差分序列的无条件期望

$$\mathrm{E}(x_i) = \mathrm{E}_{x_{i-1}, \cdots, x_1}\left[\mathrm{E}(x_i \mid x_{i-1}, \cdots, x_1)\right] = 0$$

命题　对鞅序列进行一阶差分，就得到鞅差分序列。

证明：假设$\{x_i\}_{i=1}^{\infty}=\{x_1, x_2, x_3, \cdots\}$为鞅过程。

定义其差分为$g_1 \equiv x_1$，$g_i \equiv x_i - x_{i-1}$，$\forall i \ge 2$。

对$\forall i \ge 2$，条件期望

$$
\begin{aligned}
&\mathrm{E}(g_i \mid g_{i-1}, \cdots, g_1) \\
&= \mathrm{E}(g_i \mid x_{i-1}, \cdots, x_1) && (\{g_{i-1}, \cdots, g_1\}\text{与}\{x_{i-1}, \cdots, x_1\}\text{包含同样的信息}) \\
&= \mathrm{E}(x_i - x_{i-1} \mid x_{i-1}, \cdots, x_1) && (\text{定义}\, g_i \equiv x_i - x_{i-1}) \\
&= \mathrm{E}(x_i \mid x_{i-1}, \cdots, x_1) - x_{i-1} && (\text{期望算子的线性性}) \\
&= x_{i-1} - x_{i-1} = 0 && (\text{鞅过程的定义})
\end{aligned}
$$

故$\{g_i\}_{i=1}^{\infty}$是鞅差分序列。

鞅差分序列的中心极限定理 (Central Limit Theorem for Ergodic Stationary MDS)

假设$\{\boldsymbol{g}_i\}_{i=1}^{\infty}$为渐近独立的平稳鞅差分随机向量过程，且其协方差矩阵为$\mathrm{Cov}(\boldsymbol{g}_i) = \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \boldsymbol{\Sigma}$，记$\bar{\boldsymbol{g}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i$，则

$$\sqrt{n}\,\bar{\boldsymbol{g}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$$

普通的中心极限定理仅适用于 iid 情形，此定理适用于更一般的渐近独立的平稳鞅差分序列。

5.7　大样本 OLS 的假定

假定 5.1　线性假定

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad (i = 1, \cdots, n)$$

假定 5.2　渐近独立的平稳过程(ergodic stationarity)

$(K+1)$维随机过程$\{y_i, \boldsymbol{x}_i\}$为渐近独立的平稳过程。

【例】如果样本为随机样本，则$\{y_i, \boldsymbol{x}_i\}$独立同分布，故是渐近独立的平稳过程。

假定 5.3　前定解释变量(predetermined regressors)

所有解释变量均为“前定”(predetermined)，即它们与同期的扰动项正交，即$\mathrm{E}(x_{ik}\varepsilon_i)=0$，$\forall i, k$。

由于$\mathrm{Cov}(x_{ik}, \varepsilon_i)=\mathrm{E}(x_{ik}\varepsilon_i)-\mathrm{E}(x_{ik})\mathrm{E}(\varepsilon_i)=0-0=0$，故$\boldsymbol{x}_i$与$\varepsilon_i$不相关，仿佛在$\varepsilon_i$产生之前，$\boldsymbol{x}_i$已经确定，故名“前定解释变量”。

定义如下列向量：

$$\boldsymbol{g}_i \equiv \boldsymbol{x}_i\varepsilon_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iK} \end{pmatrix}\varepsilon_i$$

则　$\mathrm{E}(\boldsymbol{g}_i)=\mathrm{E}(\boldsymbol{x}_i\varepsilon_i)=\boldsymbol{0}$。

此假定比严格外生性假定更弱，因为后者要求扰动项与过去、现在及未来的解释变量都不相关(对于时间序列数据而言)。

假定 5.4　秩条件(rank condition)

$K\times K$ 矩阵 $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ 为非退化矩阵，即其逆矩阵 $\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}$ 存在。这个条件保证了在大样本下，$(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 存在。

假定 5.5　$\boldsymbol{g}_i$ 为鞅差分序列，且其协方差矩阵 $\boldsymbol{S}\equiv\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')=\mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i')$ 为非退化矩阵。

鞅差分序列的无条件期望为 0，故假定 5.5 比假定 5.3 强。

大样本 OLS，无须假设“严格外生性”与“正态随机扰动项”，具有更大的适用性与稳健性。

5.8 OLS 的大样本性质

由于 $\boldsymbol{X} \equiv \begin{pmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}$，故 $\boldsymbol{X}'\boldsymbol{X} = (\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \cdots \ \boldsymbol{x}_n)\begin{pmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix} = \sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i'$。

其中，$\boldsymbol{x}_i \boldsymbol{x}_i'$为$K \times K$的矩阵。

定义 $\boldsymbol{S}_{XX} \equiv \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'$ (随机矩阵$\boldsymbol{x}_i\boldsymbol{x}_i'$的样本均值)。

另一方面，$\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n}\boldsymbol{x}_i y_i$。

其中，$\boldsymbol{x}_i y_i$为$K \times 1$向量。

定义 $\boldsymbol{S}_{Xy} \equiv \frac{1}{n}\boldsymbol{X}'\boldsymbol{y} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i y_i$ (随机向量$\boldsymbol{x}_i y_i$的样本均值)。因此，

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{y}}{n} = \left(\frac{\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'}{n}\right)^{-1}\left(\frac{\sum_{i=1}^{n}\boldsymbol{x}_i y_i}{n}\right) = \boldsymbol{S}_{XX}^{-1}\boldsymbol{S}_{Xy}$$

大样本理论关注当$n\to\infty$时，$\boldsymbol{S}_{XX}^{-1}$与$\boldsymbol{S}_{Xy}$的概率收敛性质。

定理(OLS 估计量的大样本性质)

(1) ($\boldsymbol{b}$为一致估计量) 在假定 5.1−5.4 之下，$\operatorname*{plim}_{n\to\infty}\boldsymbol{b}=\boldsymbol{\beta}$。

(2) ($\boldsymbol{b}$为渐近正态) 如果把假定 5.3(即$\mathrm{E}(\boldsymbol{g}_i)=\boldsymbol{0}$)强化为假定 5.5(即$\{\boldsymbol{g}_i\}$为 MDS)，则$\sqrt{n}\,(\boldsymbol{b}-\boldsymbol{\beta})\xrightarrow{d}N\left(\boldsymbol{0},\ \mathrm{Avar}(\boldsymbol{b})\right)$。

其中，$\mathrm{Avar}(\boldsymbol{b})=\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}\boldsymbol{S}\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}$，而$\boldsymbol{S}\equiv\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')=\mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i')$

(3) ($\mathrm{Avar}(\boldsymbol{b})$的一致估计量) 假设$\hat{\boldsymbol{S}}$为$\boldsymbol{S}$的一致估计量，则$\boldsymbol{S}_{XX}^{-1}\hat{\boldsymbol{S}}\boldsymbol{S}_{XX}^{-1}$是$\mathrm{Avar}(\boldsymbol{b})$的一致估计量。

证明：(1) 抽样误差可以写为

$$\boldsymbol{b}-\boldsymbol{\beta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}=\left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{\varepsilon}}{n}=\left(\frac{\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'}{n}\right)^{-1}\left(\frac{\sum_{i=1}^{n}\boldsymbol{x}_i\varepsilon_i}{n}\right)=\boldsymbol{S}_{XX}^{-1}\,\overline{\boldsymbol{g}}$$

其中，$\overline{\boldsymbol{g}}\equiv\dfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i$，$\boldsymbol{g}_i\equiv\boldsymbol{x}_i\varepsilon_i$。假定 5.2 意味着$\{\boldsymbol{x}_i\boldsymbol{x}_i'\}$也是渐近独立平稳序列，故根据渐近独立定理，$\boldsymbol{S}_{XX}\equiv\dfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\xrightarrow{p}\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$。

假定 5.4 意味着$\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}$存在，故$\boldsymbol{S}_{XX}^{-1}\xrightarrow{p}\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}$。由于$\{\boldsymbol{g}_i\equiv\boldsymbol{x}_i\varepsilon_i=\boldsymbol{x}_i(y_i-\boldsymbol{x}_i'\boldsymbol{\beta})\}$，故$\{\boldsymbol{g}_i\}$也是渐近独立平稳序列。

假定 5.3 意味着，$\bar{\boldsymbol{g}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i \xrightarrow{p} \mathrm{E}(\boldsymbol{g}_i) = \mathrm{E}(\boldsymbol{x}_i\varepsilon_i) = \boldsymbol{0}$。因此，$\boldsymbol{S}_{XX}^{-1}\bar{\boldsymbol{g}} \xrightarrow{p} \left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1} \cdot \boldsymbol{0} = \boldsymbol{0}$，故 $\operatorname*{plim}_{n\to\infty}(\boldsymbol{b}-\boldsymbol{\beta}) = \boldsymbol{0}$。

扰动项与同期解释变量不相关(假定 5.3)，是保证 OLS 一致的最重要条件。

以一元回归图示。

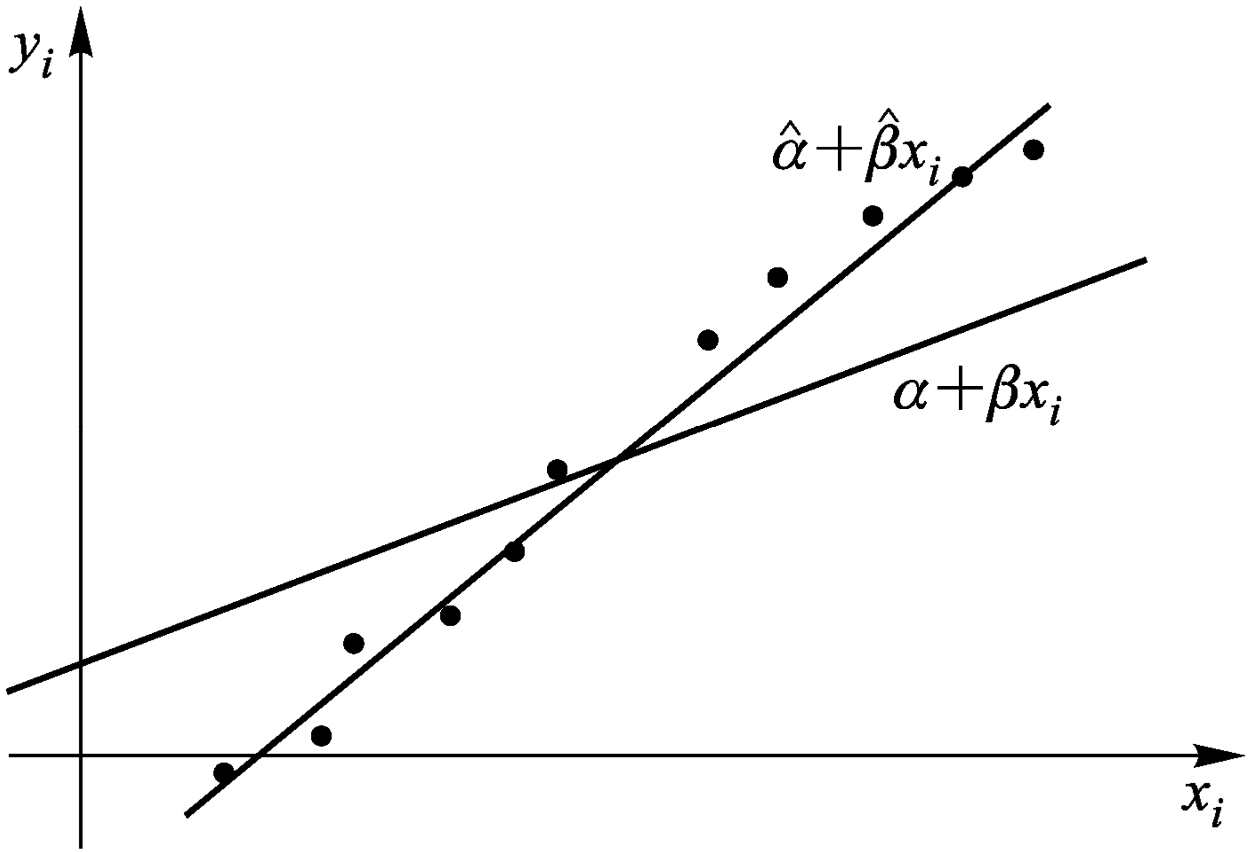

图 5.6　扰动项与解释变量相关导致不一致估计

假设$y_i = \alpha + \beta x_i + \varepsilon_i$，且$\text{Cov}(x_i, \varepsilon_i) > 0$。真实回归线$(\alpha + \beta x_i)$与样本回归线$(\hat{\alpha} + \hat{\beta} x_i)$参见图 5.6。

由于x_i与ε_i正相关，故当x_i较小时，ε_i也倾向于较小；当x_i较大时，ε_i也倾向于较大。故样本回归线比真实回归线陡峭，$\hat{\beta}$高估β。

反之，如果$\text{Cov}(x_i, \varepsilon_i) < 0$，则$\hat{\beta}$将低估$\beta$。

增大样本容量$(n \to \infty)$能使偏差(bias)消失吗？

什么情况下可能出现$\text{Cov}(x_i, \varepsilon_i) \neq 0$？

在存在遗漏变量、内生解释变量或解释变量测量误差的情况下常会出现。

(2) 由于抽样误差$\boldsymbol{b}-\boldsymbol{\beta}=\boldsymbol{S}_{XX}^{-1}\bar{\boldsymbol{g}}$，故$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})=\boldsymbol{S}_{XX}^{-1}\left(\sqrt{n}\,\bar{\boldsymbol{g}}\right)$。根据假定 5.5 及鞅差分序列中心极限定理，$\sqrt{n}\,\bar{\boldsymbol{g}}\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{S})$，其中$\boldsymbol{S}\equiv\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')=\mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i')$。

由于$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})=\boldsymbol{S}_{XX}^{-1}\left(\sqrt{n}\,\bar{\boldsymbol{g}}\right)$是$\sqrt{n}\,\bar{\boldsymbol{g}}$的线性组合，故$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})\xrightarrow{d}N\left(\boldsymbol{0},\,\mathrm{Avar}(\boldsymbol{b})\right)$。

由于$\boldsymbol{S}_{XX}\equiv\dfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\xrightarrow{p}\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$，故

$$\mathrm{Avar}(\boldsymbol{b})=\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}\boldsymbol{S}\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}$$

使用公式$\mathrm{Var}(\boldsymbol{AY})=\boldsymbol{A}\,\mathrm{Var}(\boldsymbol{Y})\boldsymbol{A}'$；其中，$\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}$为对称矩阵。

注意：无须假设扰动项服从正态分布。

(3) 如果存在 $\hat{\boldsymbol{S}} \xrightarrow{p} \boldsymbol{S}$，已知 $\boldsymbol{S}_{XX}^{-1} \xrightarrow{p} \left[\mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\right]^{-1}$，故估计量 $\widehat{\mathrm{Avar}}(\boldsymbol{b}) \equiv \boldsymbol{S}_{XX}^{-1} \hat{\boldsymbol{S}} \boldsymbol{S}_{XX}^{-1} \xrightarrow{p} \left[\mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\right]^{-1} \boldsymbol{S} \left[\mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\right]^{-1}$，是 $\mathrm{Avar}(\boldsymbol{b})$ 的一致估计量。

由于 $\boldsymbol{S}_{XX}^{-1} \hat{\boldsymbol{S}} \boldsymbol{S}_{XX}^{-1}$ 的形式为两个 $\boldsymbol{S}_{XX}^{-1}$ (“两片面包”)夹着一个 $\hat{\boldsymbol{S}}$ (“中间的菜”)，故被称为“夹心估计量”或“三明治估计量”(sandwich estimator)。

为得到 $\boldsymbol{S} \equiv \mathrm{E}(\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i')$ 的一致估计量，须对解释变量的四阶矩进行假设。

假定 5.6(解释变量的四阶矩存在) $\mathrm{E}\left[(x_{ik}x_{ij})^2\right]$存在且为有限$(\forall i, j, k)$。

这只是技术性的假定。

在假定 5.6 下，可证明$\hat{\boldsymbol{S}} \equiv \frac{1}{n}\sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$是$\boldsymbol{S}$的一致估计量，其中$\{e_i\}_{i=1}^{n}$为最小二乘法的残差。

进一步可证明，s^2是σ^2的一致估计。

命题　s^2是无条件方差$\mathrm{E}(\varepsilon_i^2)=\sigma^2$的一致估计量。

证明： $$s^2 \equiv \frac{\boldsymbol{e}'\boldsymbol{e}}{n-K}=\frac{\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}}{n-K}=\frac{\boldsymbol{\varepsilon}'\left[\boldsymbol{I}_n-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]\boldsymbol{\varepsilon}}{n-K} \quad \text{(参见第 3 章)}$$

$$=\frac{1}{n-K}\left[\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}-\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon\right] \quad \text{(乘积展开)}$$

$$=\frac{n}{n-K}\left[\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n}-\frac{\boldsymbol{\varepsilon}'\boldsymbol{X}}{n}\left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{\varepsilon}}{n}\right] \quad \text{(同时乘除 } n\text{)}$$

$$=\frac{n}{n-K}\left[\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2-\bar{\boldsymbol{g}}'\boldsymbol{S}_{XX}^{-1}\bar{\boldsymbol{g}}\right] \quad (\bar{\boldsymbol{g}}\text{ 与 }\boldsymbol{S}_{XX}\text{ 的定义})$$

由于$\operatorname*{plim}_{n\to\infty}\frac{n}{n-K}=1$，$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2=\mathrm{E}(\varepsilon_i^2)=\sigma^2$(因为$\{\varepsilon_i\}$为渐近独立的平稳序列)，$\operatorname*{plim}_{n\to\infty}\bar{\boldsymbol{g}}'\boldsymbol{S}_{XX}^{-1}\bar{\boldsymbol{g}}=\mathbf{0}'\cdot\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}\cdot\mathbf{0}=0$，故$\operatorname*{plim}_{n\to\infty}s^2=\sigma^2$。

5.9 线性假设的大样本检验

1．检验单个系数：$H_0: \beta_k = \bar{\beta}_k$。

在原假设H_0成立的情况下，$\sqrt{n}\,(b_k - \bar{\beta}_k) \xrightarrow{d} N\left(0, \mathrm{Avar}(b_k)\right)$，其中$b_k$为 OLS 估计量$\boldsymbol{b}$的第 k 个元素， $\mathrm{Avar}(b_k)$为矩阵$\mathrm{Avar}(\boldsymbol{b})$的第$(k, k)$个元素；而$\widehat{\mathrm{Avar}}(b_k)$是对$\mathrm{Avar}(b_k)$的一致估计。

定义 t 统计量：

$$t_k \equiv \frac{\sqrt{n}\,(b_k - \bar{\beta}_k)}{\sqrt{\widehat{\mathrm{Avar}}(b_k)}} = \frac{b_k - \bar{\beta}_k}{\sqrt{\frac{1}{n}\widehat{\mathrm{Avar}}(b_k)}} \equiv \frac{b_k - \bar{\beta}_k}{\mathrm{SE}^*(b_k)} \xrightarrow{d} N(0,1)$$

其中，$\mathrm{SE}^*(b_k)\equiv\sqrt{\frac{1}{n}\widehat{\mathrm{Avar}}(b_k)}=\sqrt{\frac{1}{n}\left(\boldsymbol{S}_{XX}^{-1}\hat{\boldsymbol{S}}\,\boldsymbol{S}_{XX}^{-1}\right)_{kk}}$称为“异方差稳健的标准误” (heteroskedasticity-consistent standard errors)，简称“稳健标准误”(robust standard errors，White’s standard errors，Huber-White standard errors，Eicker-Huber-White standard errors)。

在推导过程中并未用到“条件同方差”的假定，故在“条件异方差”的情况下也适用。

统计量t_k称为“稳健 t 比值”，服从标准正态分布，而非 t 分布。

$|t_k|$越大，则越倾向于拒绝H_0。比如，对于显著性水平 5%，如果$|t_k|$大于临界值 1.96，则拒绝H_0。

命题　在条件同方差的假定下，稳健标准误还原为普通(非稳健)标准误。

证明： 假设$\mathrm{E}(\varepsilon_i^2 \mid \boldsymbol{x}_i) = \sigma^2 > 0$(条件同方差)，根据迭代期望定律，

$$\boldsymbol{S} \equiv \mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i' \varepsilon_i^2) = \mathrm{E}_{\boldsymbol{x}_i} \mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i' \varepsilon_i^2 \mid \boldsymbol{x}_i) = \mathrm{E}_{\boldsymbol{x}_i} \left[\boldsymbol{x}_i \boldsymbol{x}_i' \mathrm{E}(\varepsilon_i^2 \mid \boldsymbol{x}_i) \right] = \sigma^2 \mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')$$

由于$s^2 \xrightarrow{p} \sigma^2$，$\boldsymbol{S}_{XX} \xrightarrow{p} \mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')$，故$s^2 \boldsymbol{S}_{XX}$是$\boldsymbol{S}$的一致估计量。

$$\widehat{\mathrm{Avar}}(\boldsymbol{b}) = \boldsymbol{S}_{XX}^{-1} \left(s^2 \boldsymbol{S}_{XX} \right) \boldsymbol{S}_{XX}^{-1} = s^2 \boldsymbol{S}_{XX}^{-1} = s^2 \left(\frac{1}{n} \boldsymbol{X}'\boldsymbol{X} \right)^{-1} = n s^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$\mathrm{SE}^*(b_k)=\sqrt{\frac{1}{n}\widehat{\mathrm{Avar}}(b_k)}=\sqrt{\frac{1}{n}\cdot ns^2(\boldsymbol{X'X})_{kk}^{-1}}=\sqrt{s^2(\boldsymbol{X'X})_{kk}^{-1}}$$

此公式正是“小样本 OLS”中普通(非稳健)标准误的公式。

2．检验线性假设： $H_0:\underbrace{\boldsymbol{R}}_{m\times K}\underbrace{\boldsymbol{\beta}}_{K\times 1}=\underbrace{\boldsymbol{r}}_{m\times 1}$**，其中** $\boldsymbol{R}$ **满行秩。**

根据沃尔德检验原理，考察$(\boldsymbol{Rb}-\boldsymbol{r})$的大小，譬如其二次型$(\boldsymbol{Rb}-\boldsymbol{r})'(\boldsymbol{Rb}-\boldsymbol{r})$。在$H_0$成立的情况下，统计量

$$W\equiv n(\boldsymbol{Rb}-\boldsymbol{r})'\left[\boldsymbol{R}\widehat{\mathrm{Avar}}(\boldsymbol{b})\boldsymbol{R}'\right]^{-1}(\boldsymbol{Rb}-\boldsymbol{r})\xrightarrow{d}\chi^2(m)$$

将 n 拆成$\sqrt{n}\cdot\sqrt{n}$，把 W 更直观地写为

$$W \equiv \left[\sqrt{n}(\boldsymbol{Rb}-\boldsymbol{r})\right]'\left[\boldsymbol{R}\widehat{\text{Avar}}(\boldsymbol{b})\boldsymbol{R}'\right]^{-1}\left[\sqrt{n}(\boldsymbol{Rb}-\boldsymbol{r})\right]\xrightarrow{d}\chi^2(m)$$

证明： 记$\boldsymbol{c}_n \equiv \sqrt{n}(\boldsymbol{Rb}-\boldsymbol{r})$， $\boldsymbol{Q}_n \equiv \boldsymbol{R}\widehat{\text{Avar}}(\boldsymbol{b})\boldsymbol{R}'$， 则$W = \boldsymbol{c}_n'\boldsymbol{Q}_n^{-1}\boldsymbol{c}_n$。

在H_0成立的情况下，

$$\boldsymbol{c}_n \equiv \sqrt{n}(\boldsymbol{Rb}-\boldsymbol{r}) = \sqrt{n}(\boldsymbol{Rb}-\boldsymbol{R\beta}) = \sqrt{n}\,\boldsymbol{R}(\boldsymbol{b}-\boldsymbol{\beta}) = \boldsymbol{R}\left[\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})\right]。$$

因为$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})\xrightarrow{d} N\left(\boldsymbol{0}, \text{Avar}(\boldsymbol{b})\right)$，而$\boldsymbol{c}_n$是$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})$的线性组合，故$\boldsymbol{c}_n \xrightarrow{d} \boldsymbol{c}$， 其中$\boldsymbol{c} \sim N\left(\boldsymbol{0}, \boldsymbol{R}\,\text{Avar}(\boldsymbol{b})\boldsymbol{R}'\right)$。

定义 $\boldsymbol{Q} \equiv \text{Var}(\boldsymbol{c}) = \boldsymbol{R}\,\text{Avar}(\boldsymbol{b})\boldsymbol{R}'$，由于 $\widehat{\text{Avar}}(\boldsymbol{b}) \xrightarrow{p} \text{Avar}(\boldsymbol{b})$，故 $\boldsymbol{Q}_n \xrightarrow{p} \boldsymbol{Q}$。

因此，$W = \boldsymbol{c}_n'\boldsymbol{Q}_n^{-1}\boldsymbol{c}_n \xrightarrow{d} \boldsymbol{c}'\boldsymbol{Q}^{-1}\boldsymbol{c} = \boldsymbol{c}'\left[\text{Var}(\boldsymbol{c})\right]^{-1}\boldsymbol{c} \sim \chi^2(m)$。

注： 由于$\boldsymbol{R}$满行秩，且$\text{Avar}(\boldsymbol{b})$为正定矩阵，故$\boldsymbol{Q}^{-1}$存在。