中央财经大学
Central University of Finance and Economics

计量经济学

中国财政发展协同创新中心
陈怡心
cyx@cufe.edu.cn

第4章 大样本OLS

陈怡心
中央财经大学中国财政发展协同创新中心

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趋向无穷大时统计量的性质。
- 大样本理论已成为当代计量经济学的主流方法，原因如下：
- **（1）小样本理论的假设过强。**

首先，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(不相关)。

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。

- **例** 考虑以下一阶自回归模型(first order autoregression，简记AR(1))：

$$y_t = \rho y_{t-1} + \varepsilon_t \ \ (t = 2, \cdots, T)$$

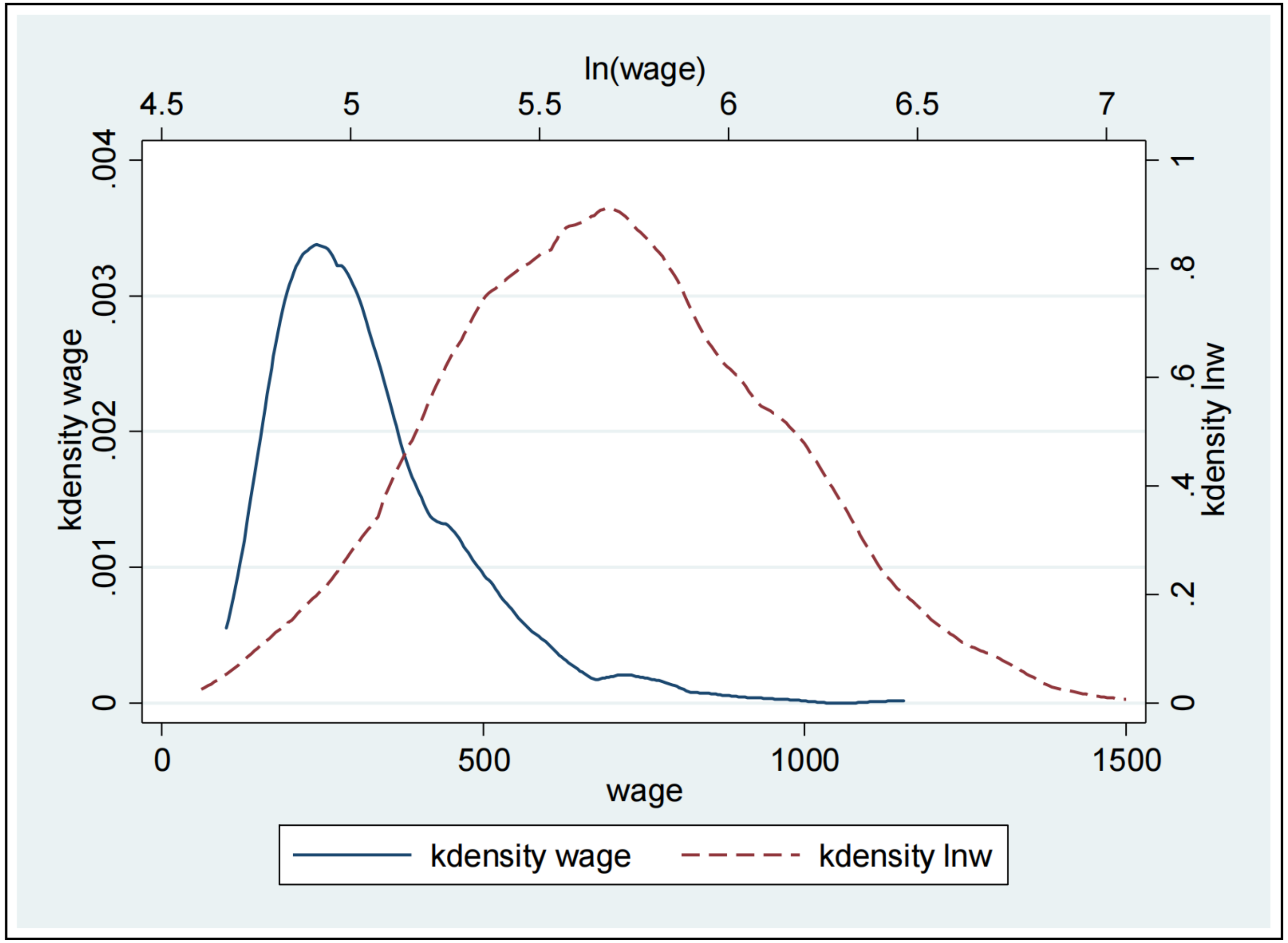
- 解释变量y_{t-1}为被解释变量y_t的一阶滞后；且$Cov(y_{t-1}, \varepsilon_t) = 0$。
- 严格外生性要求，解释变量y_{t-1}与所有$\{\varepsilon_2, \cdots, \varepsilon_T\}$均不相关。
- 这意味着，y_t也不与ε_t相关。但是ε_t是y_t的一部分，故二者一定相关，因为

$$\mathrm{Cov}(y_t, \varepsilon_t) = \mathrm{Cov}\left[(\rho y_{t-1} + \varepsilon_t), \varepsilon_t\right] = \rho \underbrace{\mathrm{Cov}(y_{t-1}, \varepsilon_t)}_{=0} + \mathrm{Var}(\varepsilon_t) = \mathrm{Var}(\varepsilon_t) > 0$$

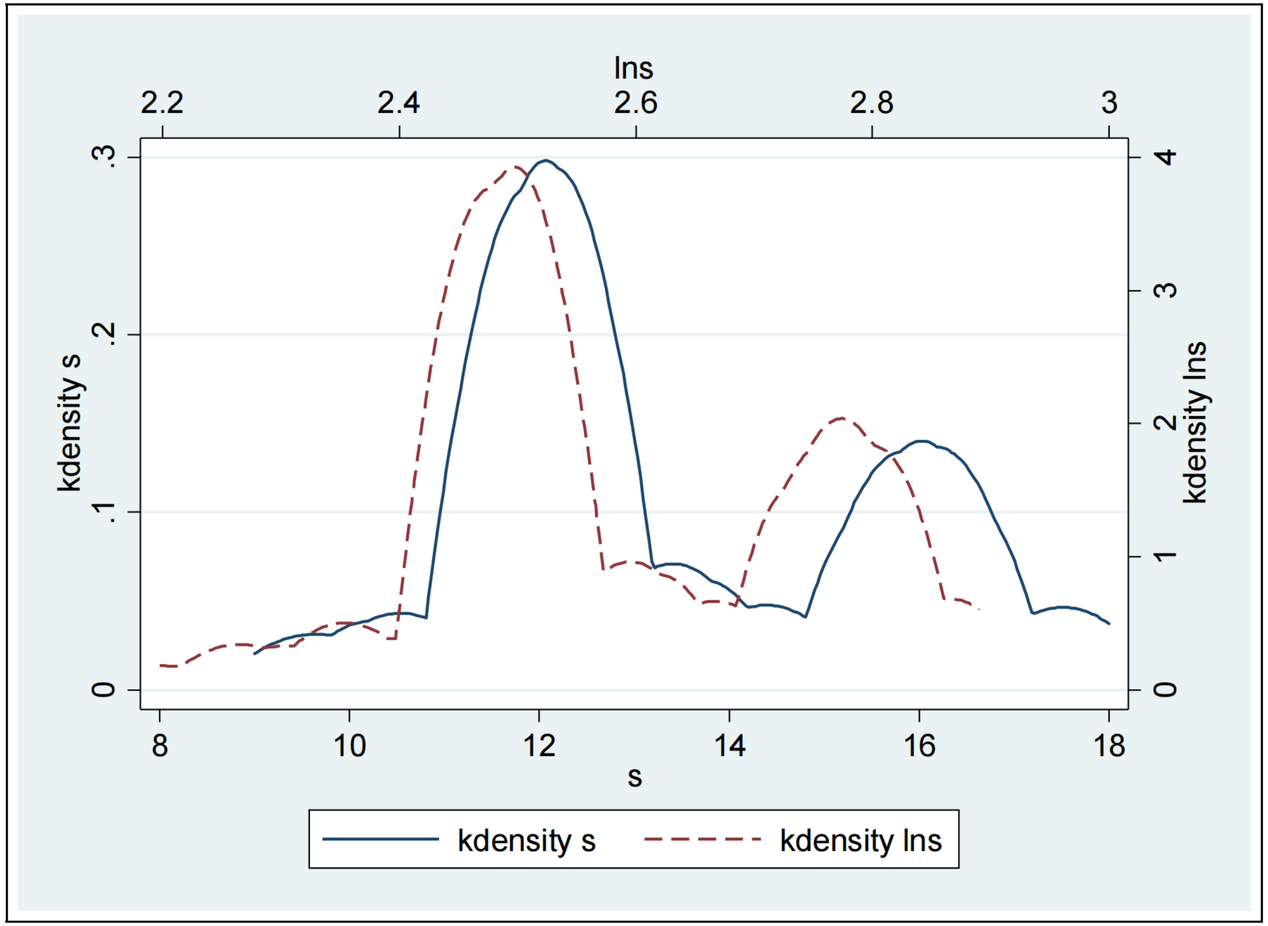
- 以被解释变量滞后值为解释变量的自回归模型，必然违背严格外生性的假定。
- **大样本理论只要求解释变量与同期(同方程)的扰动项不相关。**

- **其次，小样本理论假定扰动项为正态分布，而大样本理论无此限制。**
- 在很多情况下，并无把握经济变量是否服从正态分布。
- 比如，正态分布为对称分布，但许多经济变量的分布并不对称，例如工资收入。
- 即使考虑比较对称的工资对数，由于正态变量的取值范围为 $(-\infty,+\infty)$，而工资对数一般为正数(假设工资大于 1)，也不相符。

- 工资的分布与正态分布相去甚远。
- 工资对数，在取值范围为$(-\infty, +\infty)$上，也与正态分布不符。



- 被解释变量的分布可能为各种形状；有时即使取对数也不能使其接近正态分布。
- 教育年限的分布呈现“双峰”状，即多数人为中学或大学毕业。
- 这种双峰形状，即使取对数后，也难以改变。



- 无论教育年限还是其对数，都与“单峰”的正态分布相去甚远。
- 通过取对数使得变量的分布接近于正态并非万能。
- 对于小样本理论来说，为了进行统计推断(比如，推导t与F 统计量的分布)，须假设扰动项服从正态分布(故被解释变量也为正态)。
- 由于现实中的被解释变量可能服从各种分布(比如，变量婚否 为离散的两点分布)，故基于正态假设的小样本理论的适用范围受到很大限制。

- （2）在小样本理论的框架下，须研究统计量的精确分布(exact distribution)，但常难以推导(即使在正态分布的假设之下)。

根据大样本理论，只要研究统计量的大样本分布，即当$n \to \infty$时的渐近分布，相对容易推导(可使用大数定律与中心极限定理)。

- （3）使用大样本理论的代价是要求样本容量较大，以便大数定律与中心极限定理可以起作用。

大样本理论对于样本容量的要求，一般认为至少n>30，最好在100以上。现代的数据集越来越大，经常成百上千。

在当代计量实践中，研究人员一般用大样本理论；小样本 OLS已很少使用。

关于显著的N个tips：用大样本

- 一元线性回归下，t统计量简化为：

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})} = \frac{\hat{\beta}}{\sqrt{\hat{\sigma}^2 / nVar(x)}}$$

- 如果你的结果不显著，因为什么呢?
  —$\hat{\beta}$太小或$SE(\hat{\beta})$太大
- $\hat{\beta}$太小的潜在原因：内生性（得到的$\hat{\beta}$ 有偏）/β原本就很小
- $SE(\hat{\beta})$太大的潜在原因?
- 结果不显著，因为什么呢?
  - $\hat{\sigma}^2$很大（扰动项方差很大)
  - x的方差Var(x)很小（例如只有几个大学生)
  - 样本量n很小

为什么样本大了好?

- （1）大数定律
  - 样本不够大，发现不了规律。$\hat{\beta}$点估计值本身就不靠谱。
  - 样本小，standard error会大，导致统计显著性不过关。
  - 个别outlier的干扰会变小→影响对$\hat{\sigma}$的估计。
- （2）有足够多的花样做异质性
- （3）没有功劳，也有苦劳

数大便是美

- **徐志摩**《志摩日记二则》

数大便是美。

碧绿的山坡前几千只绵羊，挨成一片的雪绒，是美；

一天的繁星，千万只闪亮的神眼，从无极的蓝空中下窥大地，是美；

泰山顶上的云海，巨万的云峰在晨光里静定着，是美；

大海万顷的波浪，戴著各式的白帽，在日光里动荡着，是美；

爱尔兰附近的那个羽毛岛上栖息着几千万的飞禽，夕阳西沉时只见一个羽化的大空，只是万鸟齐鸣的大声，是美……

数大便是美。

数大了似乎按照著一种自然规律，自然也会有一种特别的排列，一种特别的节奏，一种特殊的式样，激动我们审美的本能，激发我们审美的情绪。

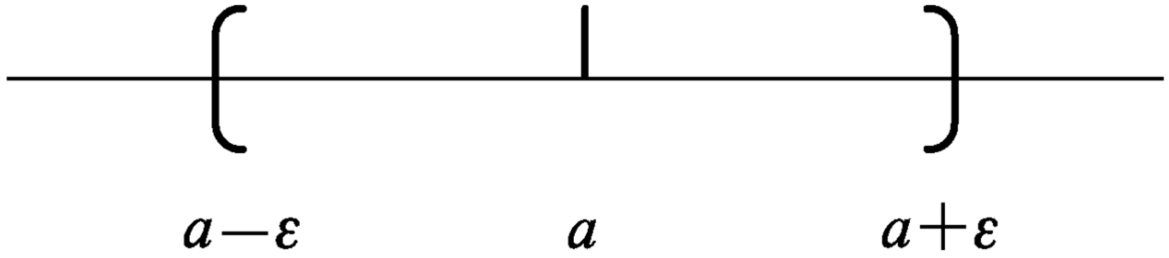
- 一些同学用分省数据、时序数据。
- 如果你的研究问题有更细的数据，就不要用粗的数据。分省VS分市VS微观

- 我们的幸运
  - 大国
  - 大数据时代

- 例外：Acemoglu, AER, The Colonial Origins of Comparative Development, 只有63个观测值!
  - 这很运气。稍微碰一下，显著性就没了。

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**（converge）于常数$a$，记为 $\lim_{n\to\infty} a_n \to a$ 或 $a_n \to a$，如果对于任意小的正数$\varepsilon > 0$，都存在$N > 0$，只要$n > N$，就有$|a_n - a| < \varepsilon$，即在a_N以后的序列$\{a_{N+1}, a_{N+2}, \cdots\}$均落入区间$(a-\varepsilon, a+\varepsilon)$内

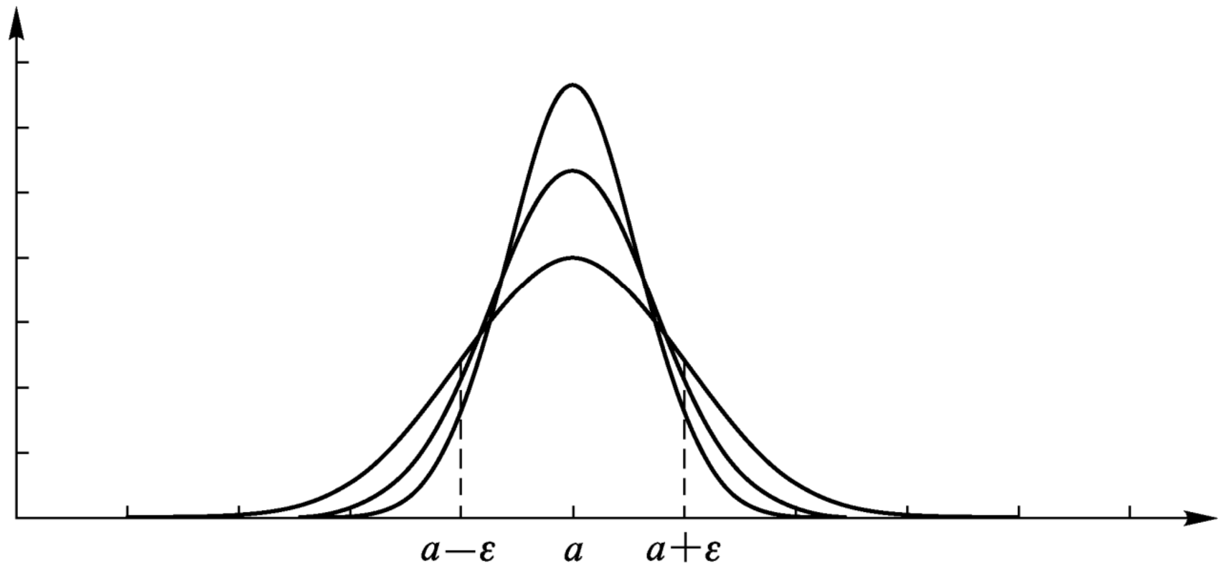


- **例** 假设$a_n = 5 + \frac{1}{n}$，则$\lim_{n\to\infty} a_n = \lim_{n\to\infty}(5 + \frac{1}{n}) = 5 + \lim_{n\to\infty}\frac{1}{n} = 5 + 0 = 5$

- **2．随机序列的收敛**
- 考虑随机序列$\{x_n\}_{n=1}^{\infty}=\{x_1,x_2,x_3,\cdots\}$，即由随机变量构成的序列，其中每个元素$x_n$都是随机变量，下标$n$通常表示样本容量。
- **定义** 随机序列$\{x_n\}_{n=1}^{\infty}$**依概率收敛**(converge in probability)于常数a，

记为$\underset{n\to\infty}{\mathrm{plim}}\, x_n=a$，或$x_n\xrightarrow{p}a$，如果对于任意$\varepsilon>0$，当$n\to\infty$时，都有$\lim\limits_{n\to\infty}P(|x_n-a|>\varepsilon)=0$。

- 任意给定很小的正数$\varepsilon>0$，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。



- 当n变大时，x_n远离常数a的可能性越来越小，变得几乎不可能。
- 由于已将随机事件$(|x_n - a| > \varepsilon)$取概率，故$P(|x_n - a| > \varepsilon)$其实是确定性序列(为概率的具体取值，已无不确定性)，而$\lim_{n\to\infty} P(|x_n - a| > \varepsilon)$只是普通的微积分极限。

- **例** 假设x_n服从如下两点分布：

$$x_n=\begin{cases}0 & \text{取值概率 } 1-(1/n)\\ n & \text{取值概率 } 1/n\end{cases}$$

- 随着$n\to\infty$， x_n的分布越来越集中于0，取值为n的可能性越来越小。故根据定义，$\operatorname*{plim}_{n\to\infty} x_n=0$。
- 利用随机变量依概率收敛于常数的概念，可定义随机变量之间的随机收敛，只要随机变量之差别依概率收敛于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。

- 概率收敛($\underset{n\to\infty}{\text{plim}}$)的运算规则类似于微积分中极限($\lim\limits_{n\to\infty}$)的运算。
- 比如，假设$g(\cdot)$为连续函数，则

$$\lim_{n\to\infty} g(x_n) = g(\underset{n\to\infty}{\text{plim}} x_n)$$

- 概率极限$\underset{n\to\infty}{\text{plim}}$与连续函数$g(\cdot)$可交换运算次序。
- 当x_n的分布越来越集中于$x^* \equiv \underset{n\to\infty}{\text{plim}} x_n$附近时，$g(x_n)$的分布自然也就越来越集中于$g(x^*)$附近。
- 期望算子$E(\cdot)$无此性质，因为$E(x^2) \neq [E(x)]^2$。这正是大样本理论的方便之处。

- **例** 如果$\operatorname*{plim}_{n\to\infty} s^2 = \sigma^2$(样本方差依概率收敛于总体方差)，则样本标准差$s$也依概率收敛于总体标准差$\sigma$，因为

$$\operatorname*{plim}_{n\to\infty} s = \operatorname*{plim}_{n\to\infty} \sqrt{s^2} = \sqrt{\operatorname*{plim}_{n\to\infty} s^2} = \sqrt{\sigma^2} = \sigma$$

- 其中，“开根号”($\sqrt{\cdot}$)是连续函数，故可与求概率极限的运算交换次序。
- 对于**随机向量序列**(即序列中每个元素都是随机向量)，也可类似地定义依概率收敛，只要定义其每个分量都依概率收敛即可。
- 比如，随机向量序列$\{\boldsymbol{x}_n\}_{n=1}^{\infty} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \cdots\}$依概率收敛于随机向量$\boldsymbol{x}$，意味着$\boldsymbol{x}$的每个分量都依概率收敛至$\boldsymbol{x}$的相应分量，记为$\operatorname*{plim}_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}$

- 3．**依均方收敛**
- **定义** 如果随机序列$\{x_n\}_{n=1}^{\infty}$的期望收敛于a，即$\lim\limits_{n\to\infty} E(x_n) = a$；而方差收敛于 0，即$\lim\limits_{n\to\infty} Var(x_n) = 0$，则称$\{x_n\}_{n=1}^{\infty}$**依均方收敛**(converge in mean square)于常数a，记为$x_n \xrightarrow{ms} a$。

$$x_1, x_2, x_3, \cdots, x_n \cdots$$

$$\mathrm{E}(\mathrm{x}_1),\mathrm{E}(\mathrm{x}_2),\mathrm{E}(\mathrm{x}_3),\cdots,E(\mathrm{x}_\mathrm{n}),\cdots \longrightarrow a$$

$$\mathrm{Var}(\mathrm{x}_1),\mathrm{Var}(\mathrm{x}_2),\mathrm{Var}(\mathrm{x}_3),\cdots,\mathrm{Var}(\mathrm{x}_\mathrm{n}),\cdots \longrightarrow 0$$

- 通过切比雪夫不等式，可以证明，依均方收敛意味着依概率收敛。

- 当x_n的均值越来越趋于a，而方差越来越小并趋于 0 时，就有$\operatorname{plim}_{n\to\infty} x_n = a$，即在极限处$x_n$退化为常数$a$。
- 证明均方收敛通常比证明概率收敛更容易，故可通过证明前者来证明后者，这也是依均方收敛概念的主要用途之一。
- 反之，依概率收敛并不意味着均方收敛。

- **例** 回到$\{x_n\}$服从两点分布的例子，即x_n取值为0的概率为$1-(1/n)$，而取值n的概率为$(1/n)$。虽然x_n依概率收敛到0，但x_n并不依均方收敛到0，因为此序列的期望恒等于1：

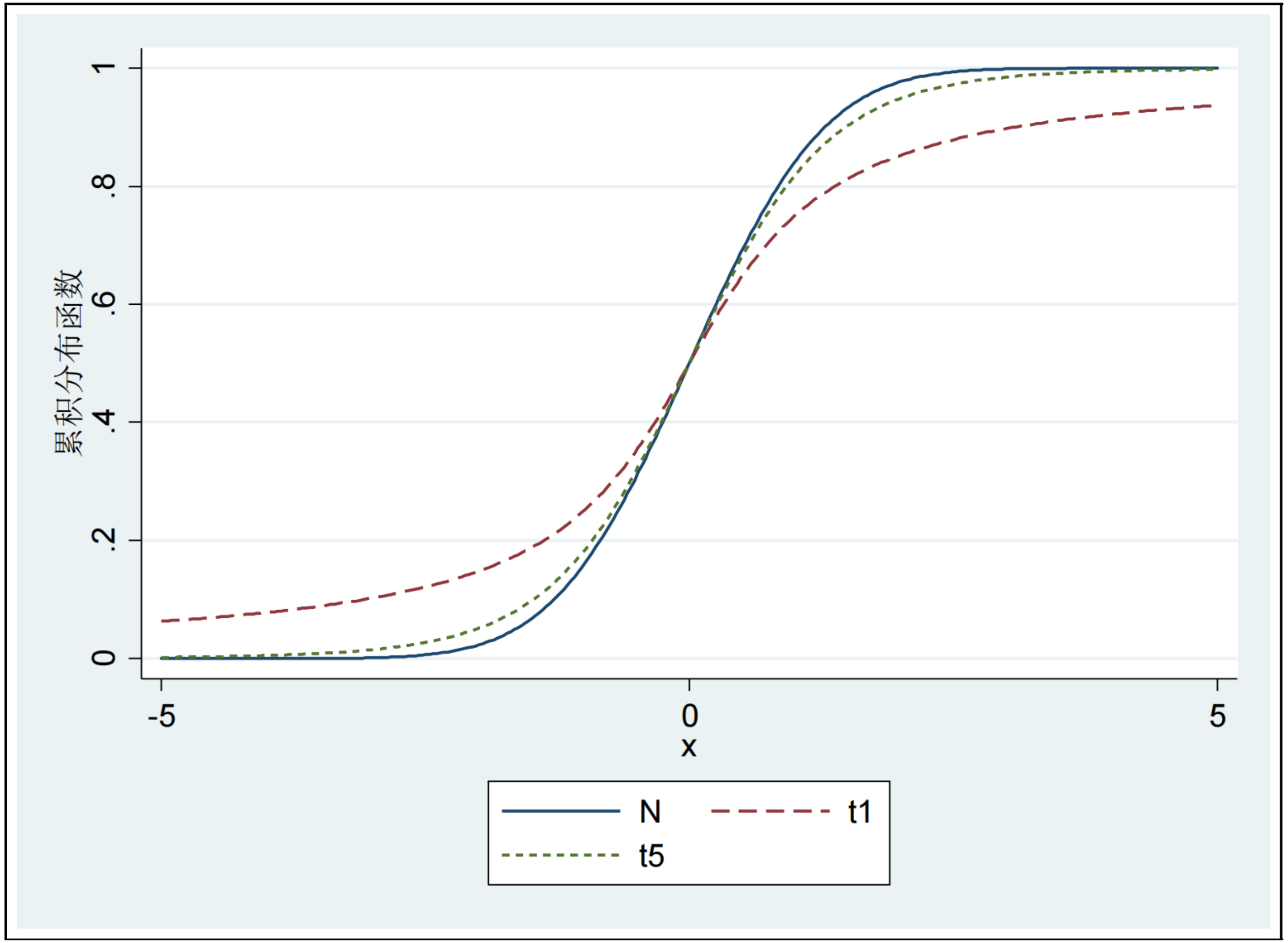
$$\lim_{n\to\infty} E(x_n) = \lim_{n\to\infty}\left[0\cdot(1-\frac{1}{n}) + n\cdot\frac{1}{n}\right] = 1 \neq 0$$

- 随着$n\to\infty$，随机序列x_n取值大于0的概率越来越小（为$1/n$），但一旦取值为正数，则很大（等于n），故此序列的期望始终为1。

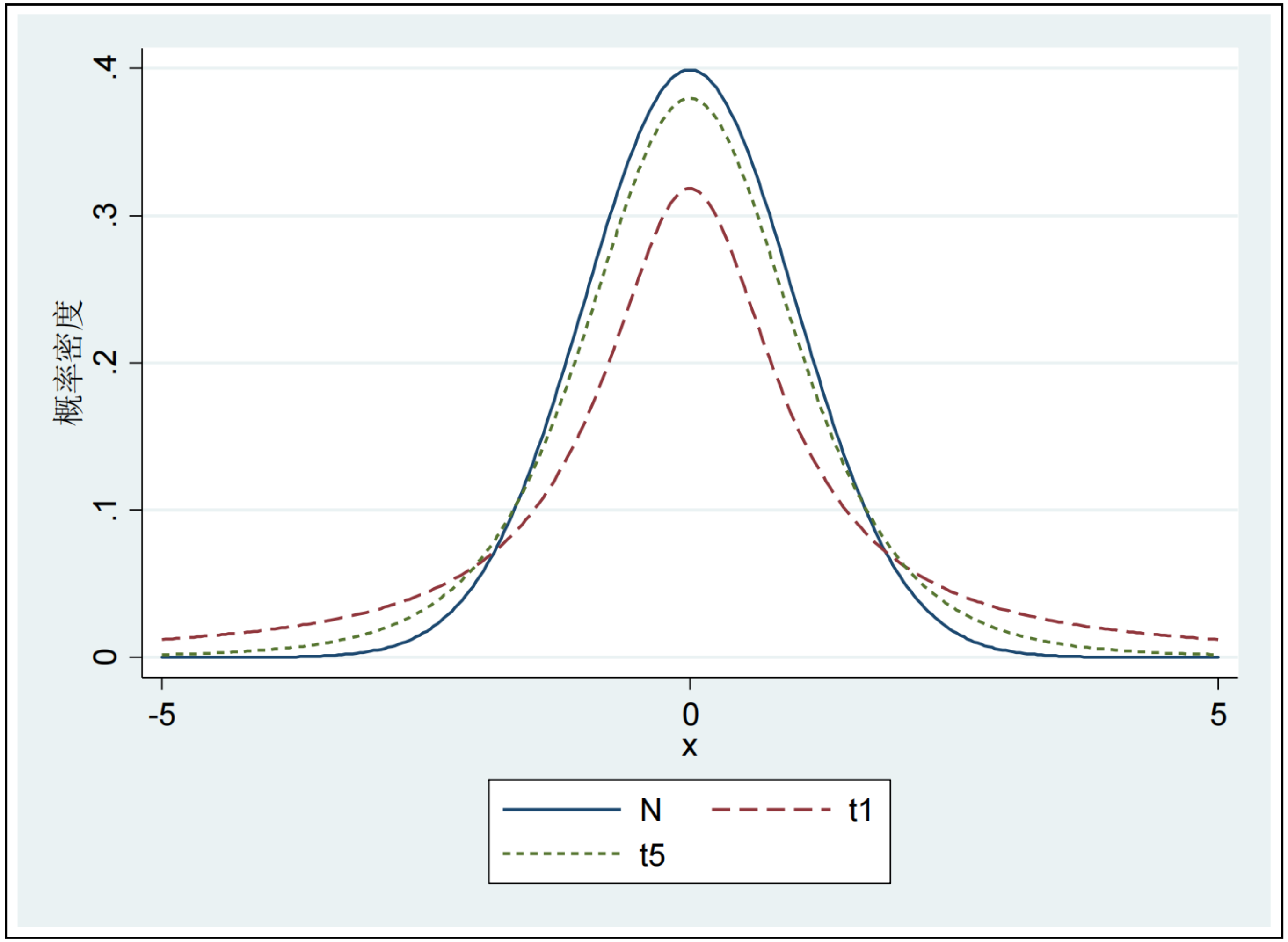
- **4．依分布收敛**
- **定义** 记随机序列$\{x_n\}_{n=1}^{\infty}$与随机变量x的累积分布函数分别为$F_n(x)$与$F(x)$。如果对于任意给定x，都有$\lim\limits_{n\to\infty} F_n(x) = F(x)$，则称随机序列$\{x_n\}_{n=1}^{\infty}$**依分布收敛**(converge in distribution)于随机变量，记为$x_n \xrightarrow{d} x$，并称x的分布为x_n的**渐近分布**(asymptotic distribution)或**极限分布**(limiting distribution)。
- 当$n \to \infty$时，x_n的分布函数越来越像x的分布函数。

- **例** 当t分布的自由度越来越大时，t分布依分布收敛于标准正态分布；即当$k \to \infty$时，$t(k) \xrightarrow{d} N(0,1)$。

- 为了直观地显示依分布收敛的过程，在 Stata 中画 N(0, 1)，t(1)与t(5)的累积分布函数



- 更直观地，可通过概率密度函数，来考察t分布依分布收敛于标准正态的过程



- 许多统计量的大样本分布均为正态分布，故引入如下概念。
- **定义**　如果$x_n \xrightarrow{d} x$，且x服从正态分布，则称x_n为**渐近正态** (asymptotically normal)，即当$n \to \infty$时，x_n的分布越来越像正态分布。
- 依分布收敛的运算也很方便。

假设$x_n \xrightarrow{d} x$，而$g(\cdot)$为连续函数，则$g(x_n)$的渐近分布就是$g(x)$，即$g(x_n) \xrightarrow{d} g(x)$。

- 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。这为大样本理论的推导提供了方便。

- **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。
- 因此，渐近标准正态的平方服从渐近$\chi(1)$的分布。

- “依概率收敛”比“依分布收敛”更强，前者是后者的充分条件；但反之，则不然。
- 如果$x_n \xrightarrow{p} x$，则意味着$(x_n - x) \xrightarrow{p} 0$，即在极限处$x_n$与$x$的具体取值无区别，故二者的概率分布也必然相同，所以$x_n \xrightarrow{d} x$。
- 如果$x_n \xrightarrow{d} x$，这只说明在极限处x_n与x的分布函数相同，但x_n与x的实际取值仍可以很不相同（比如，x_n与x相互独立）。

- **例** 假设x与y都为标准正态，且相互独立。考虑随机序列$\{x_n = x + (1/n)\}_{n=1}^{\infty}$。
- 由于$1/n \to 0$，故x_n的渐近分布为标准正态，因此$x_n \xrightarrow{d} y$（y也是标准正态）。
- 但x_n却与y相互独立，x_n的具体取值也与y毫无关系，故 x_n并不依概率收敛于y。

- 依分布收敛只是分布函数的收敛（随机变量之间可以毫无关系），而依概率收敛才是随机变量本身的收敛。
- 总之，“依均方收敛”→“依概率收敛”→“依分布收敛”。

大数定律与中心极限定理

- **1．大数定律**(Law of Large Numbers)
- 假定$\{x_n\}_{n=1}^{\infty}$为独立同分布的随机序列，且$E(x_1) = \mu$，$Var(x_1) = \sigma^2$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：首先，$E(\bar{x}_n) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n} \cdot n\mu = \mu$，故样本均值$\bar{x}_n$的期望仍为$\mu$。
- 其次，$Var(\bar{x}_n) = Var(\frac{x_1+\cdots+x_n}{n}) = \frac{1}{n^2}(n\sigma^2) \longrightarrow 0$，样本均值$\bar{x}_n$的方差收敛到0。
- 因此，$\bar{x}_n$依均方收敛于μ。
- 由此“依均方收敛”是“依概率收敛”的充分条件，故$\bar{x}_n \xrightarrow{p} \mu$。
- 当样本容量n很大时，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)
- 根据大数定律，当$n \to \infty$时，样本均值$\bar{x}_n$依概率收敛到总体均值μ。但一般情况下，$\bar{x}_n$的具体分布很难推导。
- 中心极限定理告诉我们，无论原序列$\{x_n\}_{n=1}^{\infty}$服从什么分布，当$n \to \infty$时，样本均值$\bar{x}_n$的渐近分布都为正态分布。
- 只要样本容量n足够大，则$\bar{x}_n$的真实分布将很接近于正态分布。
- **中心极限定理**　假定$\{x_n\}_{n=1}^{\infty}$为独立同分布的随机序列，且$E(x_1) = \mu$，

$Var(x_1) = \sigma^2$存在，则

$$\frac{\bar{x}_n - \mu}{\sqrt{\sigma^2/n}} \xrightarrow{d} N(0,1)$$

- 标准化之后的样本均值(即减去期望，除以标准差)的渐近分布为标准正态。

- 直观上，可视为$\bar{x}_n \xrightarrow{d} N(\mu, \sigma^2/n)$：但不严格，因为$\bar{x}_n$的方差$\sigma^2/n \to 0$(在极限处，$\bar{x}_n$的方差为 0，故退化为常数$\mu$）。
- 将表达式两边同乘σ：

$$\sigma\left(\frac{\bar{x}_n - \mu}{\sqrt{\sigma^2/n}}\right) \xrightarrow{d} \sigma N(0,1)$$

因为乘以某常数σ为连续函数，故根据依分布收敛的运算规则

$$y_n \xrightarrow{d} y \quad \Longrightarrow \quad \sigma y_n \xrightarrow{d} \sigma y$$

整理可得

$$\frac{\bar{x}_n - \mu}{\sqrt{1/n}} \xrightarrow{d} N(0, \sigma^2)$$

- 将$\sqrt{1/n}$放到分子上，可得中心极限定理的等价表达式：

$$\sqrt{n}(\bar{x}_n-\mu)\xrightarrow{d}N(0,\sigma^2)$$

其中，$\sqrt{n}\to\infty$，而根据大数定律，$(\bar{x}_n-\mu)\xrightarrow{p}0$。

故上式用$\sqrt{n}(\bar{x}_n-\mu)$（即“$\infty\cdot 0$”型）得到非退化的渐近正态分布$N(0,\sigma^2)$。

而且，$\bar{x}_n\xrightarrow{p}\mu$的速度在数量级上与$\sqrt{n}$相当，称为“root-n convergence”，即$(\bar{x}_n-\mu)\xrightarrow{p}0$的速度大约为

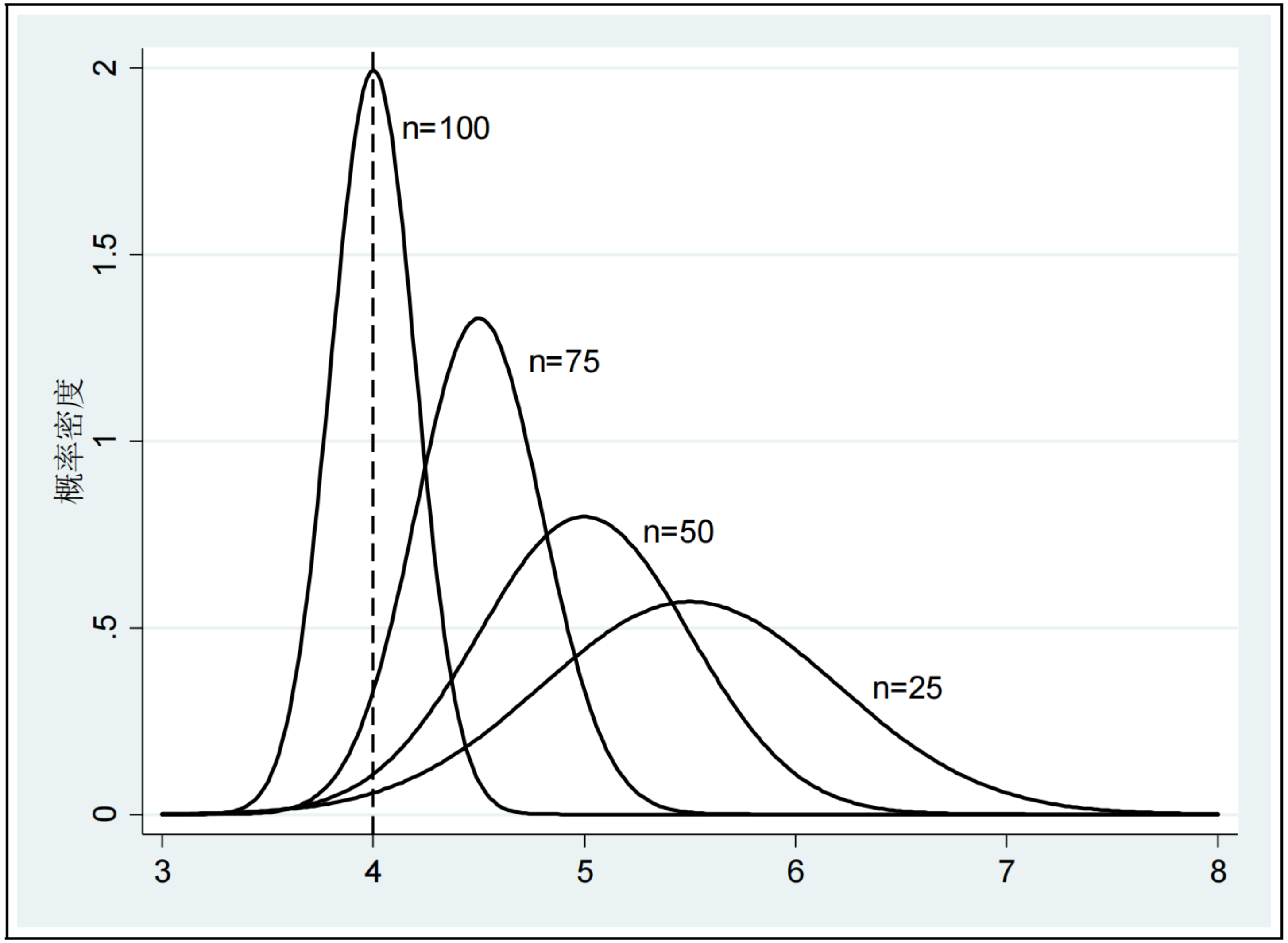
$$\frac{1}{\sqrt{1}},\frac{1}{\sqrt{2}},\frac{1}{\sqrt{3}},\cdots,\frac{1}{\sqrt{n}},\cdots$$

- 该式的好处是，容易推广到多维的情形。
- **多维的中心极限定理**：假定$\{\boldsymbol{x}_n\}_{n=1}^{\infty}$为独立同分布的随机向量序列，且$E(\boldsymbol{x}_1)=\boldsymbol{\mu}$，$Var(\boldsymbol{x}_1)=\boldsymbol{\Sigma}$存在，则$\sqrt{n}(\bar{\boldsymbol{x}}_n-\boldsymbol{\mu})\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Sigma})$。

统计量的大样本性质

- **1．一致估计量**
- **定义** 考虑参数β的估计量$\hat{\beta}_n$，其中下标n为容量（强调$\hat{\beta}_n$对样本容量n的依赖）。如果$\operatorname*{plim}_{n\to\infty}\hat{\beta}_n=\beta$，则称$\hat{\beta}_n$是参数$\beta$的一致估计量(consistent estimator)。
- 在多维情况下，称估计量$\widehat{\boldsymbol{\beta}}_n$是参数$\boldsymbol{\beta}$的一致估计量，如果$\operatorname*{plim}_{n\to\infty}\widehat{\boldsymbol{\beta}}_n=\boldsymbol{\beta}$，即$\widehat{\boldsymbol{\beta}}_n$的各分量都是$\boldsymbol{\beta}$相应分量的一致估计。
- 一致性(consistency)意味着，当样本容量足够大时，$\hat{\beta}_n$依概率收敛到真实参数β。
- 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，则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。



- **2. 渐近正态分布与渐近方差**
- 定 义　如果$\sqrt{n}(\hat{\beta}_n-\beta)\xrightarrow{d}N(0,\sigma^2)$，则称$\hat{\beta}_n$为**渐近正态**(asymptotically normal)，称σ^2为其**渐近方差**(asymptotic variance)，记为$Avar(\hat{\beta}_n)$。
- 可近似认为$\hat{\beta}_n\xrightarrow{d}N(\beta,\sigma^2/n)$，但不严格（方差$\sigma^2/n$趋于0，故为退化的分布）。
- 在多维情况下，如果 $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta})\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Sigma})$，其中 $\boldsymbol{\Sigma}$ 为半正定矩阵，则称$\widehat{\boldsymbol{\beta}}_n$为渐近正态分布，而称$\boldsymbol{\Sigma}$为$\widehat{\boldsymbol{\beta}}_n$的渐近协方差矩阵，记为$Avar(\widehat{\boldsymbol{\beta}}_n)$。

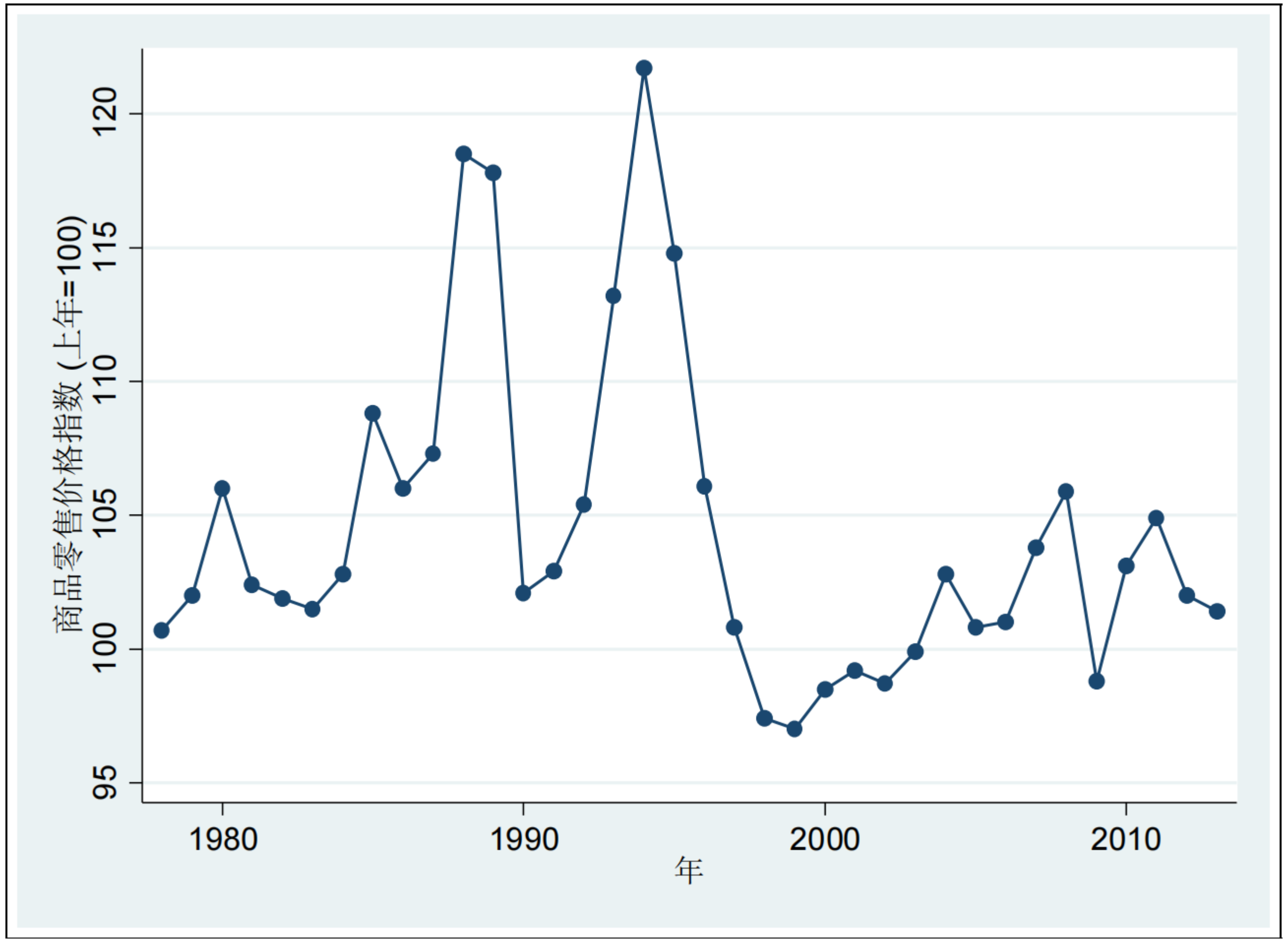
- **3. 渐近有效**
- 假设$\hat{\beta}_n$与$\tilde{\beta}_n$都是β的渐近正态估计量。如果$Avar(\hat{\beta}_n) \leq Avar(\tilde{\beta}_n)$，则称$\hat{\beta}_n$比$\tilde{\beta}_n$**更为渐近有效**(asymptotically more efficient)。
- 在大样本下，$\hat{\beta}_n$的方差小于$\tilde{\beta}_n$的方差(在小样本下未必如此)。
- 在多维情况下，假设$\widehat{\boldsymbol{\beta}}_n$与$\widetilde{\boldsymbol{\beta}}_n$都是$\boldsymbol{\beta}$的渐近正态估计量。如果$[Avar(\widetilde{\boldsymbol{\beta}}_n) - Avar(\widehat{\boldsymbol{\beta}}_n)]$为半正定矩阵，则称$\widehat{\boldsymbol{\beta}}_n$比$\widetilde{\boldsymbol{\beta}}_n$**更为渐近有效**。

随机过程的性质

- 大数定律与中心极限定理假设随机序列为 iid，但对于大多数经济变量，此假定太强。
- 比如，今年的通货膨胀率通常依赖于去年的通货膨胀率，二者并非相互独立。
- 需要研究随机序列的性质，并将推广大数定律与中心极限定理。

- 随机序列$\{x_n\}_{n=1}^{\infty}$有个更好听的名称，叫“随机过程”(stochastic process)。
- 如下标为时间，则记为 $\{x_t\}_{t=1}^{\infty}$ ，也称“时间序列”(time series)。

- 1.严格平稳过程
- 考察中国 1978—2013 年的通货膨胀率，即$\{\pi_{1978},\pi_{1979},\cdots,\pi_{2013}\}$



- 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 $E(\pi_{1978})$与$Var(\pi_{1978})$呢?
- 每年通货膨胀率的样本容量仅为 1，且历史不能重演!
- 如果这 36 年的通货膨胀率分布都不变，则可将$\bar{\pi} \equiv \frac{1}{36}\sum_{t=1978}^{2013}\pi_t$作为$E(\pi_t)$的估计量。
- 通常要求随机过程$\{x_n\}_{t=1}^{\infty}$的概率分布不随时间推移而改变。
- 无论过去、现在还是未来去看此随机过程，它的概率分布性质都一样。

- 这种随机过程称为“严格平稳过程”，它要求随机过程的有限维分布不随时间推移而改变
- 比如，相同$(\forall t, s)$;
- (x_1, x_4)的分布与(x_2, x_5)相同（二者均相隔3期）；
- (x_1, x_2, x_3)的分布与(x_5, x_6, x_7)相同（二者均为连续3期）。

- **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。
- 将$\{x_{t_1}, x_{t_2}, \cdots, x_{t_m}\}$中每个变量的时间下标全部前移或后移k期，不会改变其分布。
- $\{x_{t_1}, x_{t_2}, \cdots, x_{t_m}\}$的联合分布仅取决于$\{t_1, t_2, \cdots, t_m\}$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$\{x_t\}_{t=1}^{\infty} = \{x_1, x_1, \cdots, x_1\}$（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(AR(1)),

$$y_t = \rho y_{t-1} + \varepsilon_t \quad (t = 1, \cdots, T)$$

其中，$\{\varepsilon_t\}$为独立同分布，且$Cov(y_{t-1}, \varepsilon_t) = 0$。

- **命题** 如果$\rho = 1$，则$y_t = y_{t-1} + \varepsilon_t$。因此，$y_1 = y_0 + \varepsilon_1$，而$y_2 = y_1 + \varepsilon_2 = y_0 + \varepsilon_1 + \varepsilon_2$，以此类推可知

$$y_t = y_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$$

给定初始值y_0，当$t \to \infty$时，

$$Var(y_t) = Var(\varepsilon_1) + Var(\varepsilon_2) + \cdots + Var(\varepsilon_t) = t\sigma_\varepsilon^2 \to \infty,$$

其中$\sigma_\varepsilon^2 \equiv Var(\varepsilon_t)$,即方差越来越大，以至无穷。

- 故$\{y_t\}$不是平稳过程(平稳过程要求同分布，故方差不变)。

- 由于y_t只是在y_{t-1}的基础上，加上随机扰动项ε_t，故当$\rho = 1$时，称$\{y_t\}$为“随机游走”(random walk)。
- 如果$|\rho| < 1$，则$Var(y_t)$会收敛到常数。对方程两边同时取方差，可得

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

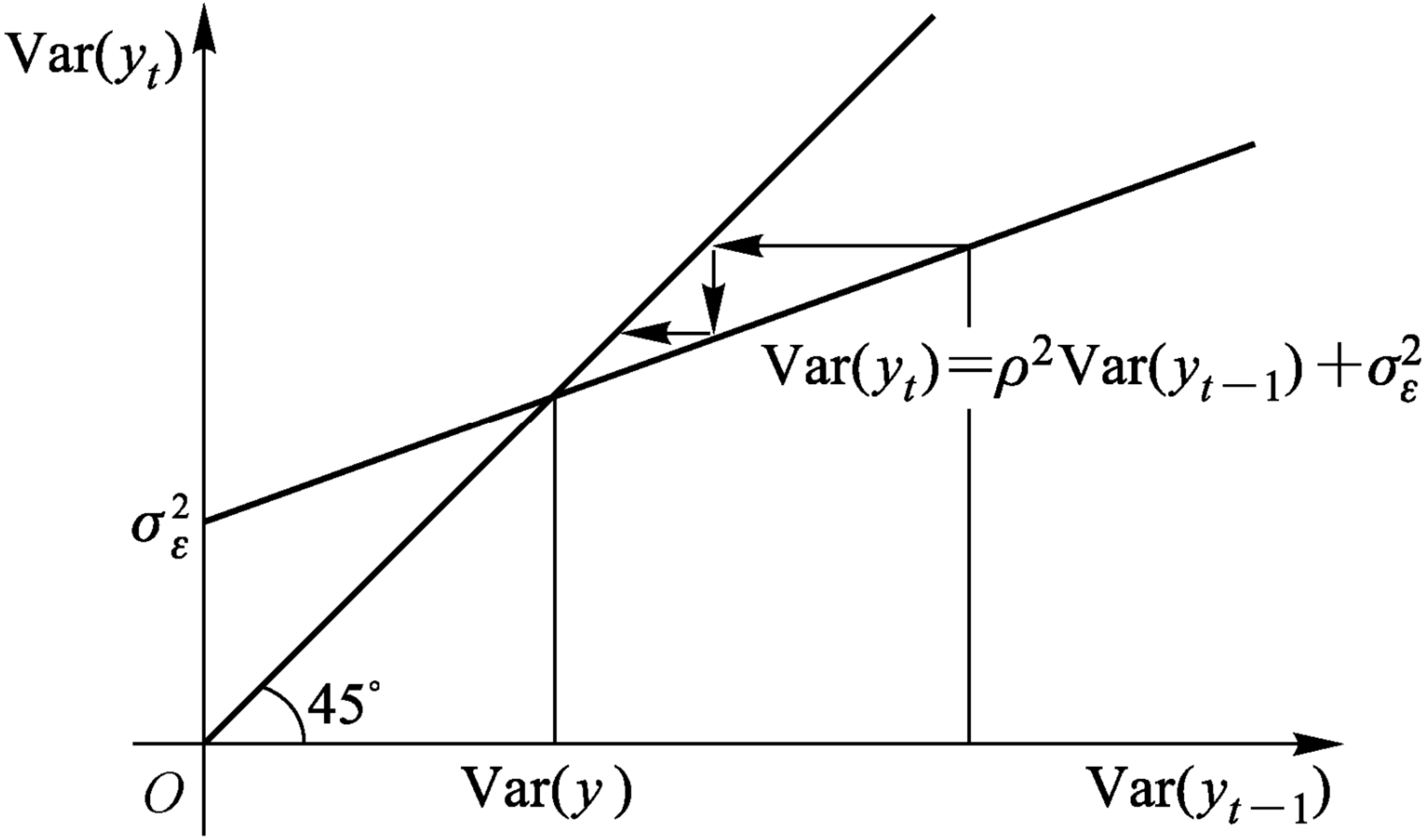
- 记$z_t \equiv Var(y_t), z_{t-1} = Var(y_{t-1})$，则上式可写为

$$z_t = \rho^2 z_{t-1} + \sigma_\varepsilon^2$$

- 这是确定性的一阶线性差分方程，因为$z_t \equiv Var(y_t)$为非随机。
- 由于$\rho^2 < 1$，故$Var(y_t)$将收敛到一个稳定值。
- 令$z_t = z_{t-1}$，可求解此收敛的稳定值z^*:

$$z^* = \rho^2 z^* + \sigma_\varepsilon^2$$

- 将上式整理后可得，$z^* = \frac{\sigma_\varepsilon^2}{1-\rho^2}$。
- 如果忽略序列$\{y_t\}$的前面几项，可将$\{y_t\}$的方差视为常数。
- 进一步可证明，$\{y_t\}_{t=0}^{\infty}$是严格平稳过程。


$\mathrm{Var}(y_t)$
$\mathrm{Var}(y_t) = \rho^2 \mathrm{Var}(y_{t-1}) + \sigma_\varepsilon^2$
σ_ε^2
45°
O
$\mathrm{Var}(y)$
$\mathrm{Var}(y_{t-1})$

- 有时仅关心随机过程的期望、方差及协方差是否稳定，而不要求整个分布都稳定，故引入以下“弱平稳过程”的概念。
- **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$E(x_t)$不依赖于t，而且$Cov(x_t, x_{t+k})$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。
- 对于弱平稳过程，由于$E(x_t)$不依赖于t，故其期望为常数。
- 由于$Cov(x_t, x_{t+k})$仅依赖于k，如果$k = 0$，则$Cov(x_t, x_t) = Var(x_t)$也不依赖于t，故弱平稳过程的方差也是常数。
- 严格平稳过程是弱平稳过程的充分条件；但反之则不然。
- 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），而概率分布还可能依赖于更高阶的矩。

- **定义** 对于弱平稳过程$\{x_t\}_{t=1}^{\infty}$，如果对于$\forall t$，都有$E(x_t)=0$，而且$Cov(x_t,x_{t+k})=0(\forall k\neq 0)$，则称为**白噪声过程**(white noise process)。
- 白噪声过程不一定独立同分布，也不一定是严格平稳过程。
- “白噪声”是性质比较好的“噪声”，即该噪声的期望值为 0，而

不同期之间的噪声互不相关。

- 推广到多维：对于随机向量过程$\{\boldsymbol{x}_t\}_{t=1}^{\infty}$，可以类似地定义平稳过程或弱平稳过程（只要将上述定义中的x置换为$\mathbf{x}$即可）。
- 对于随机向量过程$\{\mathbf{x}_t\}_{t=1}^{\infty}$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(iid)。
- 但“相互独立”的假定对于大多数经济变量过强。
- 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**)，或**弱相依**(weakly dependent)。
- 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。
- **例** 相互独立的随机序列是渐近独立的。

- **例** AR(1)是否渐近独立?

  考虑以下一阶自回归模型：

$$y_t = \rho y_{t-1} + \varepsilon_t$$

其中，$|\rho| < 1$，而ε_t为白噪声。

- 计算其各阶“自协方差”(autocovariance)。
- 当时间间隔为 1 期时，一阶自协方差为
- $Cov(y_t, y_{t-1}) = Cov(\rho y_{t-1} + \varepsilon_t, y_{t-1}) = \rho\sigma_y^2 + \underbrace{Cov(\varepsilon_t, y_{t-1})}_{=0} = \rho\sigma_y^2$
- 其中，σ_y^2为y的方差；而

$$Cov(\varepsilon_t, y_{t-1}) = Cov(\varepsilon_t, \rho y_{t-2} + \varepsilon_{t-1}) = Cov(\varepsilon_t, \varepsilon_{t-1}) = 0$$

因为ε_t为白噪声。

- 当时间间隔为2期时，原方程可写为

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

- 因此，二阶自协方差为
- $Cov(y_t, y_{t-2}) = Cov(\rho^2 y_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t, y_{t-2}) = \rho^2\sigma_y^2$

- 以此类推，当时间间隔为j期时，

$$Cov(y_t, y_{t-j}) = \rho^j\sigma_y^2$$

- 由于$|\rho| < 1$，故当上式$j \to \infty$时，$Cov(y_t, y_{t-j}) \to 0$。
- 相距越远，则序列$\{y_t\}$的自协方差越小，且在极限处变为 0(不相关)，故此 AR(1)模型为渐近独立的过程。

- **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$E(x_i)=\mu$存在，则$\bar{x}_n \equiv \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{p} \mu$，即样本均值$\bar{x}_n$是总体均值$E(x_i)$的一致估计。
- 渐近独立定理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即可。
- 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，故也是同分布的。
- 类似地，可将中心极限定理作相应的推广；即在一定条件下，中心极限定理也适用于渐近独立的平稳过程。

- **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$f(\cdot)$，$\{y_i \equiv f(x_i)\}_{i=1}^{\infty}$也是渐近独立的严格平稳过程。

- 根据此命题，则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)$E[f(x_i)]$，都可以由其对应的“样本矩”(sample moment)$\frac{1}{n}\sum_{i=1}^{n} f(x_i)$来一致地估计。

- **例** 对于渐近独立的平稳过程 $\{x_i\}_{i=1}^{\infty}$ 样本方差 $s^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$，是总体方差$Var(x) = E[x - E(x)]^2$的一致估计。
- **例** 样本协方差$s_{xy} \equiv \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$，为总体协方差$Cov(x, y) \equiv E[(x - E(x))(y - E(y))]$的一致估计。

大样本OLS的假定

- **假定1** 线性假定

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \ (i = 1, \cdots, n)$$

此假定与小样本OLS完全相同。

- **假定2** $(K+1)$维随机过程$\{y_i, x_{i1}, \cdots, x_{iK}\}$为渐近独立的平稳过程(ergodic stationarity)，故适用大数定律与中心极限定理。
- **例** 如果样本为随机样本，则$\{y_i, x_{i1}, \cdots, x_{iK}\}$独立同分布，故是渐近独立的平稳过程。

- **假定3** 前定解释变量(predetermined regressors)

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

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

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)，而前定变量仅要求与同期的扰动项不相关。

- **假定4** 秩条件(rank condition)

数据矩阵$\boldsymbol{X}$ 满列秩，即$\boldsymbol{X}$ 中没有多余(可由其他变量线性表出)的解释变量，故不存在严格多重共线性。

- 大样本理论的假定 1 与 4 与小样本理论相同，而假定 2 与3 则比小样本理论更为放松。
- 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。

OLS的大样本性质

- 在假定1-4之下，OLS 估计量$\hat{\boldsymbol{\beta}}$具有以下良好的大样本性质。
- （1）**$\hat{\boldsymbol{\beta}}$为一致估计量**，**即**$\underset{n\to\infty}{\text{plim}}\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$

  以一元回归为例。考虑以下模型：

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, \cdots, n)$$

β的OLS估计量为

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- 此模型的离差形式为

$$(y_i-\bar{y})=\beta(x_i-\bar{x})+(\varepsilon_i-\bar{\varepsilon})$$

$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$， $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$， $\bar{\varepsilon}=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i$

- 代入$\hat{\beta}$的表达式可得

$$\hat{\beta}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})[\beta(x_i-\bar{x})+(\varepsilon_i-\bar{\varepsilon})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$=\frac{\beta\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(x_i-\bar{x})(\varepsilon_i-\bar{\varepsilon})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$=\beta+\frac{\sum_{i=1}^{n}(x_i-\bar{x})(\varepsilon_i-\bar{\varepsilon})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$=\beta+\frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(\varepsilon_i-\bar{\varepsilon})}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}\xrightarrow{p}\beta+\underbrace{\frac{\mathrm{Cov}(x_i,\varepsilon_i)}{\mathrm{Var}(x_i)}}_{=0}=\beta$$

其中，$Cov(x_i,\varepsilon_i)=0$

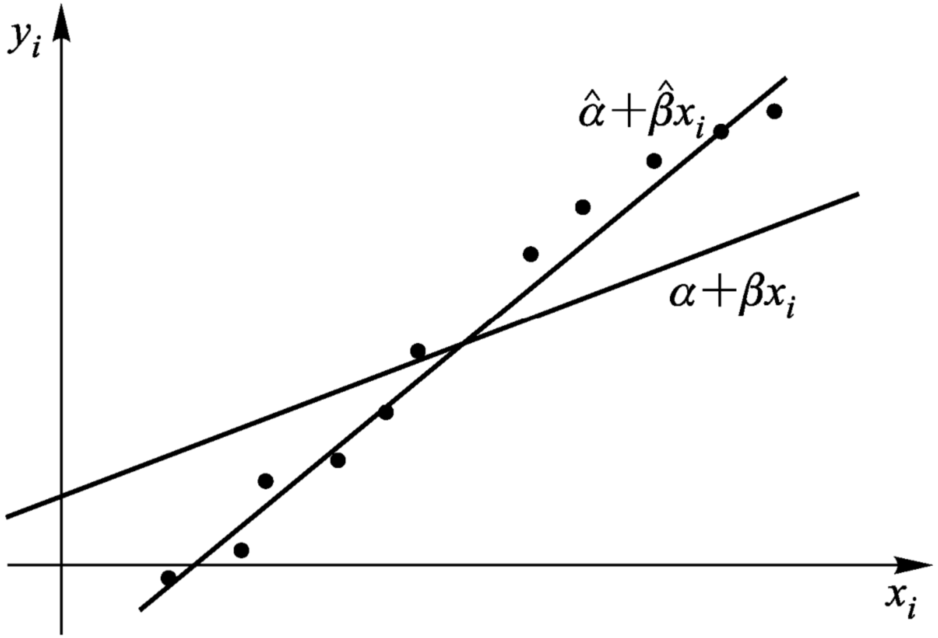
- 前定解释变量，或扰动项与解释变量同期不相关，是保证 OLS一致的最重要条件。
- 反之，如果$Cov(x_i, \varepsilon_i) \neq 0$，，则$\underset{n\to\infty}{\text{plim}}\hat{\beta} = \beta + \frac{Cov(x_i,\varepsilon_i)}{Var(x_i)} \neq \beta$
- 如果$Cov(x_i, \varepsilon_i) > 0$，则$\underset{n\to\infty}{\text{plim}}\hat{\beta} > \beta$

　　比如，考察教育投资的回报率，x_i为教育年限，而ε_i为被遗漏的个人能力。x_i与ε_i正相关(能力高者通常上学更久)，故 OLS 估计量将高估教育投资的回报率。

- 如果$Cov(x_i, \varepsilon_i) < 0$，则$\underset{n\to\infty}{\text{plim}}\hat{\beta} < \beta$

　　比如，考察上医院对健康的作用，x_i为是否上医院，而ε_i为个人原来的健康状况(被遗漏)。x_i与ε_i负相关(通常只有健康不佳者才上医院)，故 OLS 估计量将低估上医院对健康的正面作用(去医院者的健康往往不如未去医院者)。

- 通过图示考察$Cov(x_i, \varepsilon_i) \neq 0$的后果
- 真实(总体)回归线为$\alpha + \beta x_i$
- 样本回归线为$\hat{\alpha} + \hat{\beta} x_i$
- 假设$Cov(x_i, \varepsilon_i) > 0$，$x_i$与$\varepsilon_i$正相关
- 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
- 反之，如果$Cov(x_i, \varepsilon_i) < 0$，则$\hat{\beta}$将低估$\beta$。
- 增大样本容量($n \to \infty$)也不能使偏差(bias)消失。



- 在计量经济学中，如果解释变量与扰动项相关，即$Cov(x_i, \varepsilon_i) \neq 0$，则称此解释变量为“内生解释变量”(endogenous regressor)，简称“内生变量”。反之，则为“外生变量”(exogenous variable)。
- 由于内生变量的存在，致使 OLS 回归出现偏差，统称为“内生性偏差”(endogeneity bias)，或简称“内生性”。

- 在什么情况下可能出现内生性偏差?
- 如果存在**遗漏变量**、**双向因果关系**、或**解释变量测量误差**

(measurement errors)，常会出现解释变量与扰动项同期相关的情形，导致 OLS 不一致。

- (2) $\hat{\boldsymbol{\beta}}$**服从渐近正态分布，即**$\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\xrightarrow{d}N\left(\mathbf{0},Avar(\hat{\boldsymbol{\beta}})\right)$

  其中，$Avar(\hat{\boldsymbol{\beta}})$为$\hat{\boldsymbol{\beta}}$的渐近协方差矩阵。

  $\hat{\boldsymbol{\beta}}$之所以服从渐近正态，因为在一定条件下，中心极限定理适用于渐近独立的平稳过程。

- (3) 由于大样本理论一般不假设球形扰动项，故渐近协方差矩阵 $Avar(\hat{\boldsymbol{\beta}})$的表达式更为复杂。
- OLS 估计量$\hat{\boldsymbol{\beta}}$的协方差矩阵可写为：

$$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

其中，$Var(\boldsymbol{\varepsilon}|\boldsymbol{X})$为扰动项的协方差矩阵。

- 如果存在球形扰动项(同方差、无自相关)，则$Var(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2\boldsymbol{I}_n$
- 上式简化为

$$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\sigma^2\boldsymbol{I}_n)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 对于横截面数据，经常存在异方差，但无自相关(比如，各截面单位之间相互独立)。
- 考虑存在条件异方差，但无自相关的情形。
- 扰动项的协方差矩阵可写为

$$\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}$$

- 其中，$\sigma_1^2, \cdots, \sigma_n^2$不全相等。

- 如何估计上式的$\{\sigma_1^2, \cdots \sigma_n^2\}$?
- 其中，$\sigma_1^2 = Var(\varepsilon_1) = E(\varepsilon_1^2) - [E(\varepsilon_1)]^2 = E(\varepsilon_1^2)$
- 以 OLS 残差平方$\{e_1^2, \cdots e_n^2\}$替代上式的$\{\sigma_1^2, \cdots \sigma_n^2\}$，得到扰动项协方差矩阵的估计量：

$$\widehat{\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})} = \frac{n}{n-K} \begin{pmatrix} e_1^2 & & 0 \\ & \ddots & \\ 0 & & e_n^2 \end{pmatrix}$$

- 其中，$\frac{n}{n-K}$为自由度的调整（在大样本下无差别）。

- 将上式代入$Var(\hat{\boldsymbol{\beta}}|\boldsymbol{X})$的表达式，可得如下方差估计量

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 考虑$\sqrt{n}\hat{\boldsymbol{\beta}}$的方差估计量，即$\hat{\boldsymbol{\beta}}$的**渐近方差估计量**：

$$\widehat{\mathrm{Avar}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})} = n(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 上式为$\hat{\boldsymbol{\beta}}$渐近协方差矩阵的一致估计量，即

$$\widehat{\mathrm{Avar}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})} \xrightarrow{p} \mathrm{Avar}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})$$

- 由于表达式在推导过程中并未假设“条件同方差”，故在“条件异方差”情况下也成立，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)。

- 在形式上，稳健标准误也是夹心估计量。
- 稳健标准误的思想最早由 Eicker(1967)与 Huber(1967)提出，并由 White(1980) 严格证明，故也称 White’s standard errors， Huber-White standard errors，或 Eicker-Huber-White standard errors。
- 稳健标准误的表达式虽较复杂，但对于计算机，其计算成本可以忽略(无须人为记忆)。
- 通过使用迭代期望定律可以证明，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。

- 考虑同方差的极端情形，即$e_1^2=e_2^2=\cdots=e_n^2$(所有残差的绝对值都相等，但符号可以相反),则

$$\widehat{\mathrm{Var}(\boldsymbol{\varepsilon}\mid\boldsymbol{X})}=\frac{n}{n-K}\begin{pmatrix} e_1^2 & & 0 \\ & \ddots & \\ 0 & & e_n^2 \end{pmatrix}=\frac{ne_i^2}{n-K}\boldsymbol{I}_n=\frac{\sum_{i=1}^{n}e_i^2}{\underbrace{n-K}_{=s^2}}\boldsymbol{I}_n=s^2\boldsymbol{I}_n$$

- 稳健的协方差矩阵简化为同方差情况下的普通(非稳健)协方差矩阵：

$$\begin{aligned}\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}\mid\boldsymbol{X})}&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\mathrm{Var}(\boldsymbol{\varepsilon}\mid\boldsymbol{X})}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\\&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(s^2\boldsymbol{I}_n)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}=s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\end{aligned}$$

大样本统计推断

- 对于渐近独立的平稳过程，如果样本容量足够大，则 OLS 估计量 $\hat{\boldsymbol{\beta}}$的渐近正态分布是对其真实分布的较好近似，可使用其渐近分布进行大样本假设检验与区间估计。
- 大样本统计推断(large sample inference)的步骤与小样本 OLS 基本相同。

- **1．检验单个系数：**H_0：$\beta_k = c$
- 考虑检验H_0：$\beta_k = c$，其中c为已知常数。
- 根据大样本理论，OLS 估计量$\hat{\boldsymbol{\beta}}$服从渐近正态分布，即$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, Avar(\hat{\boldsymbol{\beta}}))$，其中$Avar(\hat{\boldsymbol{\beta}})$为渐近协方差矩阵。
- 具体到$\hat{\boldsymbol{\beta}}$的第k个元素$\hat{\beta}_k$，则有

$$\sqrt{n}(\hat{\beta}_k - \beta_k) \xrightarrow{d} N(0, Avar(\hat{\beta}_k))$$

- $Avar(\hat{\beta}_k)$为$\hat{\beta}_k$的渐近方差，即渐近方差矩阵$Avar(\hat{\boldsymbol{\beta}})$主对角线上的第$k$个元素。

- 在原假设H_0成立的情况下，$\beta_k = c$，故表达式可写为

$$\sqrt{n}(\hat{\beta}_k - c) \xrightarrow{d} N(0, Avar(\hat{\beta}_k))$$

- 记$\widehat{Avar(\hat{\beta}_k)}$为渐近方差矩阵估计量$\widehat{Avar(\hat{\boldsymbol{\beta}})}$主对角线上的第$k$个元素，则$\widehat{Avar(\hat{\beta}_k)}$是$Avar(\hat{\beta}_k)$的一致估计量。
- 定义t统计量为

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

- $SE^*(\hat{\beta}_k) \equiv \sqrt{\frac{1}{n}\widehat{Avar(\hat{\beta}_k)}}$即为异方差稳健的标准误。
- 统计量t_k称为“稳健t比值”(robust t ratio)，服从渐近标准正态分布，而不是t分布。

- 对于双边检验(即H_1：$\beta_k \neq c$)，则$|t_k|$越大，越倾向于拒绝H_0。
- 比如，对于 5%的显著性水平，如果$|t_k|$大于临界值1.96，则可拒绝H_0。
- 也可以通过p值进行检验，方法与小样本理论相同。

- **2．检验线性假设：**H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$
- 考虑检验m个线性假设是否同时成立：

$$H_0: \underbrace{\boldsymbol{R}}_{m\times K} \underbrace{\boldsymbol{\beta}}_{K\times 1} = \underbrace{\boldsymbol{r}}_{m\times 1}$$

- 其中，$\boldsymbol{r}$为m维列向量$(m < K)$,$\boldsymbol{R}$为$m \times K$矩阵。
- $rank(\boldsymbol{R}) = m$，即$\boldsymbol{R}$满行秩，没有多余或自相矛盾的行或方程。
- 对于原假设H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$ ，根据沃尔德检验原理，可考察$(\boldsymbol{R\hat{\beta}} - \boldsymbol{r})$的大小，譬如其二次型$(\boldsymbol{R\hat{\beta}} - \boldsymbol{r})'(\boldsymbol{R\hat{\beta}} - \boldsymbol{r})$。

- 在H_0成立的情况下，可证明统计量

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

- $\boldsymbol{R}Avar(\hat{\boldsymbol{\beta}})\boldsymbol{R}'$为$(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})$的渐近方差矩阵(使用夹心估计量公式)。
- 如果统计量W大于$\chi^2(m)$的临界值，则拒绝原假设。
- 虽然统计量W服从χ^2分布，而非小样本的F分布，但χ^2分布与F分布在大样本情况下是等价的。
- 即使在大样本下使用稳健标准误进行假设检验，Stata 也依然汇报F统计量及其 p值。

- **命题** 假设统计量$F \sim F(m,n)$分布，则当$n \to \infty$时，$mF \xrightarrow{d} \chi^2(m)$。
- **证明**：因为$F \sim F(m,n)$，故可写为$\mathrm{F} = \frac{\chi^2(m)/m}{\chi^2(n)/n}$，其中分子与分母相互独立。
- 根据χ^2分布的性质，χ^2分布的期望等于自由度，而方差等于自由度的两倍；即$E[\chi^2(n)] = n$，且$Var[\chi^2(n)] = 2n$。
- 分母的期望为$E[\chi^2(n)/n] = n/n = 1$，而方差为$\mathrm{Var}[\chi^2(n)/n] = 2n/n^2 = 2/n \longrightarrow 0$(当$n \longrightarrow \infty$时)。
- 分母依均方收敛于 1，故依概率收敛于 1(前者是后者的充分条件)，即$\chi^2(n)/n \xrightarrow{p} 1$。
- F统计量的性质仅由分子$\chi^2(m)/m$决定，故$F \xrightarrow{d} \chi^2(m)/m$。
- 因此，在大样本下，$mF \xrightarrow{d} \chi^2(m)$。

一元回归实例

```
. reg lnw s
```

	Source	SS	df	MS
ESS	Model	35.2039946	1	35.2039946
RSS	Residual	104.082155	756	.137674809
TSS	Total	139.28615	757	.183997556

Number of obs	=	758	
F(1, 756)	=	255.70	
Prob > F	=	0.0000	
R-squared	=	0.2527	R^2
Adj R-squared	=	0.2518	
Root MSE	=	.37105	

		回归系数	标准误				
被解释变量	lnw	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
解释变量	s	.0966245	.0060425	15.99	0.000	.0847624	.1084866
常数项	_cons	4.391486	.0821136	53.48	0.000	4.230288	4.552684

$$\hat{\alpha} = 4.391，\ \hat{\beta} = 0.097$$

$$\widehat{\ln w} = 4.391 + 0.097s$$

教育的投资回报率为9.7%，
每增加一年教育，平均可提高收入9.7%

多元回归实例

```
. reg lnw s expr tenure smsa rns
```

	Source	SS	df	MS
ESS	Model	49.0478814	5	9.80957628
RSS	Residual	90.2382684	752	.119997697
TSS	Total	139.28615	757	.183997556

Number of obs =	758	
F(5, 752) =	81.75	F统计量
Prob > F =	0.0000	P值
R-squared =	0.3521	R^2
Adj R-squared =	0.3478	Adjusted R^2
Root MSE =	.34641	

		回归系数	标准误	t统计量	P值	置信区间	
被解释变量	lnw	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
解释变量	s	.102643	.0058488	17.55	0.000	.0911611	.114125
	expr	.0381189	.0063268	6.02	0.000	.0256986	.0505392
	tenure	.0356146	.0077424	4.60	0.000	.0204153	.0508138
	smsa	.1396666	.0280821	4.97	0.000	.0845379	.1947954
	rns	-.0840797	.0287973	-2.92	0.004	-.1406124	-.0275471
	_cons	4.103675	.085097	48.22	0.000	3.936619	4.270731

```
. esttab , se r2 b(4)  star(* 0.1 ** 0.05  *** 0.01) nogap
```

	(1)
	lnw
s	0.1026***
	(0.0062)
expr	0.0381***
	(0.0066)
tenure	0.0356***
	(0.0080)
smsa	0.1397***
	(0.0281)
rns	-0.0841***
	(0.0295)
_cons	4.1037***
	(0.0877)
N	758
R-sq	0.352

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

```
.  reg lnw s expr tenure smsa rns, robust
```

Linear regression

Number of obs	=	758
F(5, 752)	=	84.05
Prob > F	=	0.0000
R-squared	=	0.3521
Root MSE	=	.34641

lnw	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
s	.102643	.0062099	16.53	0.000	.0904523	.1148338
expr	.0381189	.0066144	5.76	0.000	.025134	.0511038
tenure	.0356146	.0079988	4.45	0.000	.0199118	.0513173
smsa	.1396666	.028056	4.98	0.000	.0845893	.194744
rns	-.0840797	.029533	-2.85	0.005	-.1420566	-.0261029
_cons	4.103675	.0876665	46.81	0.000	3.931575	4.275775

谢谢！