计量经济学笔记

ElegantLaTeX 版本

作者：滑翔闪

组织：ElegantLaTeX Program

时间：December 16, 2025

ElegantLaTeX Program

因果推断是一门哲学

目录

前言

All econometrics are about two things: Law of Large Number (LLN) and Central Limit Theorem (CLT)

— –Yongmiao Hong

我觉得老师说得对，其实在国外没有初级、中级、高级计量之分，区别在于将统计数学思想和细节领悟到多细。即便统计学学了一遍假设检验和推断统计，计量学了最小二乘的基本性质，概率论学了分布和推断统计，但上课时还是疯狂遗忘，归根结底我没在一个框架内理解这些知识。

一个有趣的悖论是——思想在简单数学中就可以领悟了 (微观经济学其实最是如此)，但是大部分人需要学会更复杂的数学语言后才能快速领悟，或者在返璞归真中感受到中级模型的简洁之美————这就是有没有一个好老师的区别——在简单数学语言时就埋下理解的伏笔。

对于所谓的三高，真正的难度取决于老师的水平和要求，而不是课程的名称，所以这份笔记我只称呼为计量经济学笔记而不是冠以中级或者高级之名。虽然都强调自学，但我个人感受是，无论老师水平如何，课程要求如何，上过三高课程后，我自己才真正能开始自学——重要的是在课程中认识到这个体系的框架，并且在练习中适应相关的语言表述。

本科时基于《为什么》明白了反事实框架。曾经以为第一层理解是数学拟合，只是下层；第二层是反事实的设计。如今才知道，光是数学统计的拟合，其思想也深不见底。我现在才知道, 为什么很多人说《基本无害的计量经济学》降低了计量的门槛。它更像是一个实验手册，指导人们怎么设计实验。但是计量中还有更多的统计原理和思想。基本无害的框架强调因果效应而回避了不少统计本家思想细节的引入和介绍。

学习目标，推荐参考史震涛老师的知乎回答——经济学专业博士如何学好高级计量？-史震涛的回答。

学习过程：现在互联网资源很丰富，数学类的逻辑，不懂的原因都可以概括为一类——**知识过渡的不自然**，好在如今的互联网资源数不胜数，多谷歌即可补充理解的润滑油。不懂可以抄书，也是我整理这份笔记的原因，虚构观众并整理笔记，至少能知道自己是哪里不懂。还有一点，这种笔记可以反复使用，以后可以继续补充。

Ai 时代，代码使用和理解之间可以毫不相关，同样的，知道和理解一个知识之间总隔着一层自我傲慢的幻觉。如何证明自己真的学懂了？如何证明自己自学是有效的？我一向认为还是需要外界的一点点助力。

学习信仰：抱着怀疑前进，勇于思考学习。计量诞生之初是为了实现预测，但预测终究是一个极其复杂的综合问题。如今，计量被看作讲经济寓言故事的工具。有人说计量摧残了经济思想，也有人说计量使得经济学更为科学。经济学家究竟是在进行福尔摩斯般环环相扣的推导？还是进行东方锦衣卫那样苦苦相逼的数据拷打？这些都是值得在学习中思考的问题。计量是在说服一个人相信一个规律，还是试图证明一个真实存在的规律，这其中的差距才是做学问的差距。

因果分析不存在金科玉律，计量只是一种思考框架。经济增长理论模型和因果推断计量模型实在是个人觉得经济学门槛最高，但也最奇妙的研究方向。卢卡斯说，当你开始思考起经济增长，你就很难再思考其他问题，因果推断也是如此。什么是因果？如何用数据刻画一件事的作用？如何证明一个规律确实存在？在实证分析中，概率多大才算大？多少的数据证据才能算得上证明？

以上问题似乎没有完美的答案，却塑造了因果推断永恒的魅力。对于因果，即便是上帝之力，恐怕也不外乎如此了。

第一章 概率论与数理统计预备知识：渐近理论

1.1 数列的收敛

定义 1.1

确定数列有界的常用表示：对于数列 $\{a_N\}$，如果 $N^{-\lambda}a_N$ 有界，我们记 $a_N = O(N^{\lambda})$。特别地，当 $\{a_N\}$ 本身有界时,$a_N = O(1)$。

确定数列有极限的常用表示：对于数列 $\{a_N\}$，如果 $N^{-\lambda}a_N \to 0$，我们记 $a_N = o(N^{\lambda})$。特别地，当 $\{a_N\}$ 本身有极限时,$a_N = o(1)$。

随机数列将以上换为 x

引理 1.1

引理：假如 $w_N = o_p(1), x_N = o_p(1), y_N = O_p(1), z_N = O_p(1)$。那么，下面等式成立：

$$(1) w_N + x_N = o_p(1)$$
$$(2) y_N + z_N = O_p(1)$$
$$(3) y_N * z_N = O_p(1)$$
$$(4) x_N * z_N = o_p(1)$$

总结：“+”同小得小，有大得大；“*”同大得大，有小得小。

1.2 随机向量或矩阵的收敛

定义 1.2

随机向量极限的定义：设 $\{\boldsymbol{x}_N\}$ 为 $K\times 1$ 维随机向量，假如 $x_N j \xrightarrow{p} a_j (j = 1, 2, \cdots, K)$。那么，称 $x_N \xrightarrow{p} \boldsymbol{a}$。

随机矩阵极限的定义：设 $\{\mathbf{Z}_N\}$ 为 $M\times K$ 维随机矩阵，假如 $z_{ij} \xrightarrow{p} b_{ij} (i = 1, 2, \cdots, M; j = 1, 2, \cdots, K)$。那么，称 $\mathbf{Z}_N \xrightarrow{p} \boldsymbol{B}$。其中,$b_i j$ 是矩阵 B 的元素。

总结：随机向量或随机矩阵收敛就是其中的每个元素收敛。

1.3 随机向量或矩阵收敛的一些性质

定理 1.1 (斯勒茨基定理)

(Slutsky's Theorem): 设 $g: R^K \to R^J$ 是定义域 $c \in R^K$ 上的连续函数，随机向量 x_N 满足 $x_N \xrightarrow{p} \boldsymbol{c}$。那么

$$plimg(x_N) = g(plimx_N) = g(c)$$

总结：连续函数的概率极限等于概率极限的函数。

定理 1.2 (随机矩阵的可逆性)

对于随机方阵 $\{\mathbf{Z}_N\}(K\times K)$, 若存在可逆方阵 A 使得 $\mathbf{Z}_N\xrightarrow{p}\boldsymbol{A}$。那么,

$\mathbf{Z}_N^{-1}$ exists w.p.a.l(with probability approaching one)

$\mathbf{Z}_N^{-1}\xrightarrow{p}\boldsymbol{A}^{-1}$, 或者 $plim\mathbf{Z}_N^{-1}=\boldsymbol{A}^{-1}$

总结：随机方阵 $\mathbf{Z}_N$ 可以是奇异的, 但随着 $n\to\infty$,$\mathbf{Z}_N$ 奇异的概率趋于 0, 这就不会影响渐近分析。

定理 1.3 (随机矩阵的可逆性)

对于随机方阵 $\{\mathbf{Z}_N\}(K\times K)$, 若存在可逆方阵 A 使得 $\mathbf{Z}_N\xrightarrow{p}\boldsymbol{A}$。那么,

$\mathbf{Z}_N^{-1}$ exists w.p.a.l(with probability approaching one)

$\mathbf{Z}_N^{-1}\xrightarrow{p}\boldsymbol{A}^{-1}$, 或者 $plim\mathbf{Z}_N^{-1}=\boldsymbol{A}^{-1}$

总结：随机方阵 $\mathbf{Z}_N$ 可以是奇异的, 但随着 $n\to\infty$,$\mathbf{Z}_N$ 奇异的概率趋于 0, 这就不会影响渐近分析。

定理 1.4 (随机数列分布收敛)

设 $\{x_N\}$ 为随机数列,x 是连续随机变量,F_N 为随机变量 x_N 的累积分布函数 (c.d.f.),F 是连续随机变量 x 的累积分布函数。那么, 对于任意 $\xi\in R$, 当且仅当

$$F_N(\xi)\to F(\xi),N\to\infty$$

称随机数列 $\{x_N\}$ 分布收敛于 x, 记为 $x_N\xrightarrow{d}x$ 或 $x_N\overset{a}{\sim}x$。

定理 1.5 (随机向量分布收敛)

设 $\{x_N\}$ 为 $K\times 1$ 维随机向量, $\boldsymbol{x}$ 为连续随机向量, 当且仅当

$$\forall K\times 1\text{维非随机向量}c\text{满足}c'c=1,c'x_N\to c'x$$

称随机向量 $\{x_N\}$ 分布收敛于 x, 记为 $x_N\xrightarrow{d}x$ 或 $x_N\overset{a}{\sim}x$。

1.4 随机向量分布收敛的性质

引理 1.2 (分布收敛必有界)

$$x_N\to x\to x_N=O_p(1)$$

引理 1.3 (连续映射定理 (Continuous Mapping Theorem))

假如随机向量 $\boldsymbol{x}_N\xrightarrow{d}\boldsymbol{x}$, g:$R^K\to R^J$ 是连续函数, 那么,

$$g(\boldsymbol{x}_N)\xrightarrow{d}g(\boldsymbol{x})$$

总结：连续函数的概率分布等于概率分布的函数。

引理 1.4 (渐近等价定理 (Asymptotic Equivalence Lemma))

对于随机向量 x_N 和 $\mathbf{z}_N\xrightarrow{d}\mathbf{z}$, 并且 $\boldsymbol{x}_N-z_N\xrightarrow{p}0$, 那么,

$$x_N\xrightarrow{d}z$$

1.5 随机向量分布收敛的性质

引理 1.5 (分布收敛的简单运算)

假如 $K \times 1$ 维随机向量 $\{\mathbf{z}_N\}$

满足 $\mathbf{z}_N \xrightarrow{d} Normal(\mathbf{0}, \boldsymbol{v})$, 那么

对于任意非随机矩阵 $\boldsymbol{A}_{K \times M}, \boldsymbol{A}' z_N \xrightarrow{d} Normal(\mathbf{0}, \boldsymbol{A}' V \boldsymbol{A})$

$\mathbf{z}_n' \boldsymbol{V}^{-1} \mathbf{Z}_N \xrightarrow{d} \chi_K^2$

总结：上述第一条性质类似于数乘一个随机变量，第二条性质可以简单看作正态分布标准化变化的平方。

引理 1.6 (依概率收敛必然依分布收敛)

若 $x_N \xrightarrow{p} x$, 那么,$x_N \xrightarrow{d} x$, 即 $\lim_{N \to \infty} F_N(x) = F(x)$ (Slutsky's Theorem)

1.6 渐进性

关于渐进性的讲义——计量经济学讲义

基于样本趋于无穷时的特点来研究。

注[渐进性的检验]

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{d} Normal(0, \boldsymbol{V})$$

V 是 $\sqrt{N}(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$ 的渐近方差 (随机变量的方差在样本量趋近于无穷时的极限值), 可记为 $Avar\sqrt{N}(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) = V$

Wald 检验： Wald 检验是一类检验方法

若 $\sqrt{N}(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{d} Normal(0, \boldsymbol{V})$, 对于任意非随机矩阵, 有 $R_{Q \times P}, Rank(R) = Q$,

$$\sqrt{N}\boldsymbol{R}\left(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right) \xrightarrow{d} Normal\left(0, \boldsymbol{R}\boldsymbol{V}\boldsymbol{R}'\right)$$
$$\left[\sqrt{N}\boldsymbol{R}\left(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right)\right]' [\boldsymbol{R}\boldsymbol{V}\boldsymbol{R}']^{-\mathbf{1}}\left[\sqrt{N}\boldsymbol{R}\left(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right)\right] \xrightarrow{d} \chi_Q^2$$

Delta method：

推荐参考义——Delta method

$$\sqrt{N}\left[\boldsymbol{c}\left(\widehat{\boldsymbol{\theta}}_N\right) - \boldsymbol{c}(\boldsymbol{\theta})\right] \xrightarrow{d} Normal(0, \boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{V}\boldsymbol{C}(\boldsymbol{\theta})')$$
$$\left\{\sqrt{N}\left[\boldsymbol{c}\left(\widehat{\boldsymbol{\theta}}_N\right) - \boldsymbol{c}(\boldsymbol{\theta})\right]\right\}' [\boldsymbol{R}\boldsymbol{V}\boldsymbol{R}']^{-1}\left\{\sqrt{N}\left[\boldsymbol{c}\left(\widehat{\boldsymbol{\theta}}_N\right) - \boldsymbol{c}(\boldsymbol{\theta})\right]\right\} \xrightarrow{d} \chi_Q^2$$

定义 $C(\theta) = \nabla_\theta c(\theta)$ 为函数　的　$\times$　维雅可比矩阵 (Jacobian)(每一行使用相同的函数, 每一列使用相同的变量, 求偏导)。

1.7 线性代数相关知识

矩阵的转置、逆矩阵、矩阵的 (半) 正定与 (半) 负定、幂等矩阵、矩阵的秩、矩阵的迹。

第二章 简单 OLS

2.1 研究对象

$$Econometrics\begin{cases}\hat{\beta}\\ SE(\beta)\end{cases}\quad \text{估计}\begin{cases}OLS\\ GMM\\ MLE\end{cases}\quad \text{检验}\begin{cases}Wald\\ LM\\ LR\end{cases}\quad \text{类型}\begin{cases}\text{数据}\begin{cases}time\\ panel\\ APC\end{cases}\\ \begin{cases}\text{参数, 如 OLS}\\ \text{非参数, 如机器学习}\end{cases}\end{cases}$$

- **随机现象**：单次无规律, 多次有规律
- **随机试验**：条件重复, 结果范围已知, 实验前具体结果未知
- **样本空间**：随机实验的所有可能结果的集合
- **样本点**：样本空间的一个点
- **随机变量**：样本空间上对应的实值函数。总体样本在被抽取之前，其中的任意个体是随机变量，但是总体样本的值并非随机变量。在计量方程中，X_i、Y_i 为随机变量。
- **参数**：参数是总体分布的特征性常数，它本身不是随机变量。例如方程中的 β。
- **估计量**：样本特征 (非随机变量), 一般通过公式表示。总体样本的估计量不是随机变量。
- **估计值**：估计量的具体值, 例如 $\hat{\beta}_1$=2.5
- **抽样误差**: 一般为均值的误差。
- **总体**: 全体个体或观测值的集合，是总体单位而不是变量集合。例如研究 xx 大学的身高，xx 大学全体学生才是总体而非身高集合。

例题 2.1 估计量和估计值 残差没有经济含义，只代表真实值和测量值的差异，误差有经济含义。

例题 2.2 总体回归方程只有一个，样本回归方程有多个（抽样样本不同，方程不同），但是形式要和总体方程一致。

定义 2.1

什么是线性，就参数而言是线性的。对参数求导后不依赖于参数。线性是相对于于参数 β_0 而言的概念。

♣

例题 2.3 线性于参数 对于式子 $y = \beta_0 + \sqrt{\beta_1}x + u$(非线性)。如果估计试图通过 $\beta^* = \sqrt{\beta_1}$ 的方式进行转化，最终也不正确，因为此时分布的假设不同。前者估计量最后参数应该平方，此时服从的是卡方分布（估计量不能比较）。(比较 $\hat{\beta}_1^*$ 与 $\hat{\beta}_1^2$，分布一开始就不同)：更直观的例子是对一个抛物线进行线性拟合，自然是不合适的。

我们在描述一个**总体方**程关系时，总体参数 β 并不是随机变量。

$$y = \beta x + u$$

但我们在描述一个**估计方程**时，这时候 $\hat{\beta}$ 是随机变量，并且是估计量，其具体值为估计值。

OLS 是均方误估计, 倒过来最好理解, 误差平方和最小——min MSE

如果分布正确, 极大似然估计法最有效。

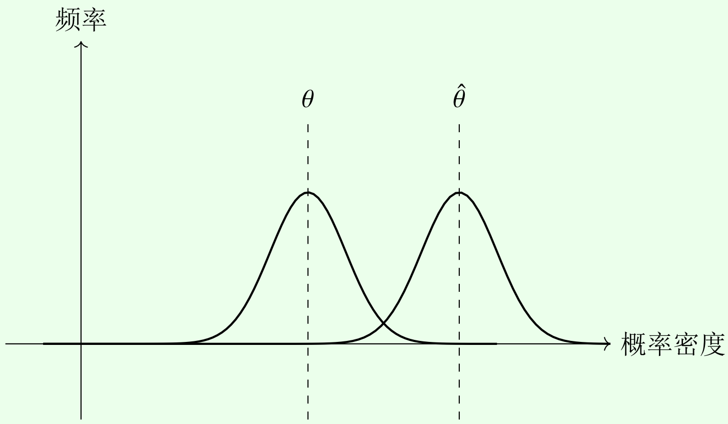
标准误是标准差的估计量。

置信区间：重复抽样,n% 的区间包含确定值。

2.2 希望性质

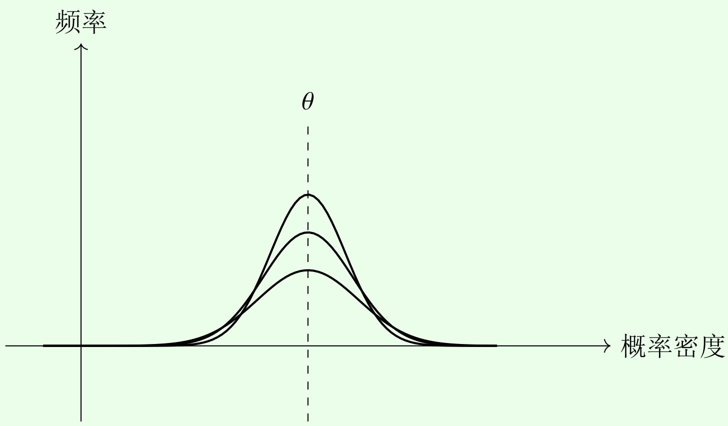
无偏性：重复抽样, 相同数量, 不同样本。

估计值等于期望值。注意，期望为样本分布的期望。所以这里重点是重复抽样。

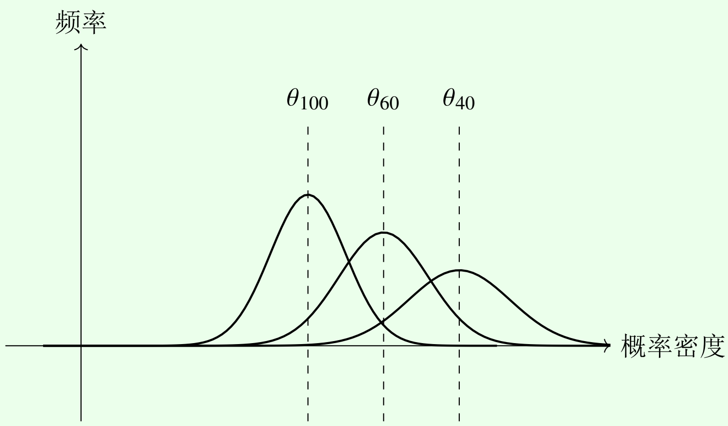


有效性：不同方法, 相同样本, 相同数量, 方差更小。

估计量的方差越小——偏离真是值的差异越小越好。



一致性：相同方法, 样本容量增多



问题 2.1 一致性和无偏性有关系吗？-无关。

无偏性：分布的差异。（样本量固定时）

一致性：渐进性的差异。（样本量可变）

总体方差和样本方差比较, 直接使用总体方差的公式, 一致但不无偏。

可参考为什么样本方差 (sample variance) 的分母是 n-1？使用 n 的情况有偏但一致。

均值 vs 渐进性，x/n，无偏但不一致。

2.3 简单最小二乘

2.3.1 相关性分析到最小二乘

先来看看**相关性**分析的式子：

$$\bar{x} = n^{-1}\sum_{i=1}^{n} x_i; \quad \bar{y} = n^{-1}\sum_{i=1}^{n} y_i$$

$$s_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2; \quad s_y^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$s_{xy} = (n-1)^{-1}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})$$

s_{xy}（协方差）的直觉理解为：减去均值的变化方向。相同方向则正，反之则反。但是直接使用 s_{xy}，单位放缩会影响大小，因此加入分子减少单位变化的影响。于是我们进一步分析有了相关性分析。

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}} = Cor(x,y)$$

。

注意：相关性分析的式子为：$\frac{s_{xy}}{s_x s_y}$

定理 2.1

柯西不等式（Cauchy-Schwarz inequality）证明了 r_{xy} 范围在-1 到 1 之间

柯西不等式：

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \le \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)$$

注 什么是自由度：其实就是样本随机变动的约束。例如 n 个数据，已知均值，当给定 n-1 个数据的值时，第 n 个样本值其实就已经确定了。在样本方差中，n-1 中的-1 部分损失的就是均值对应的自由度。

注 **自由度与坦克公式的联系**

自由度的直观含义可以通过“坦克问题”进一步理解。该问题出现在二战时期，盟军根据缴获的德军坦克编号来推算其总产量。设敌方坦克的最大编号为 N（即总数量），我们从中抽取 n 辆，观察到的最大编号记为 m。

此时，我们得到的信息是”敌人至少有 m 辆坦克”，而不是”敌人有 m 辆坦克”。所以有一种说法就是我国红军编号会跳数字就是为了扰乱敌方判断。当然，更多原因其实是军队内部调整的历史原因。

如果直接用比例 n/m 来估计总数 N，就会系统性地低估。因为样本中最大值 m 总是小于等于 N，并且随着样本量增加才逐渐逼近 N。这种低估偏差，与样本方差直接用 n 做分母时的低估本质一致。

基于极大似然或无偏性推导，可得著名的 坦克公式：

$$\hat{N} = m + \frac{m}{n} - 1.$$

其中：

- m 表示样本中观测到的最大编号；
- n 为样本量；
- $\hat{N}$ 为估计的总坦克数。

这个修正项 $\frac{m}{n} - 1$ 正好弥补了“低估”的系统偏差。比如例子中缴获了 $n = 4$ 辆，编号分别为 $1, 3, 4, 7$，则样本最大值 $m = 7$。代入公式可得：

$$\hat{N} = 7 + \frac{7}{4} - 1 = 7 + 1.75 - 1 = 7.75 \approx 8.$$

即我们会合理地估计敌方大约有 8 辆坦克，而不是简单的 7 辆。

因此，无论是样本方差的分母 $n - 1$，还是坦克问题中的修正公式，其背后共同的思想是：**有限样本信息必然丢失一个“自由度”，直接用样本统计量估计总体会产生偏差，需要进行修正。**

2.3.2 均方误最小

为什么计量比起中位数、众数、分位数，选择了条件期望，因为条件期望是预测检测上最有效的。条件期望是一种集中趋势。一般衡量模型好坏的思路是以下式子：

$$\min E(y - g(y))^2$$

此时 $g(y) = E(y)$ 是最优点

对于

$$y = \beta_0 + \beta_1 x + u$$

均方误最小————有

$$SSR(\hat{\beta}_0, \hat{\beta}_1) = min \sum_i^n u^2$$

分别对 $\hat{\beta}_0, \hat{\beta}_1$ 求导研究一阶条件，可以得到一般式子

$$\frac{\partial SSR}{\partial \beta_0} = 2 \sum (y - \beta_1 x - \beta_0) = 0$$

移项，除以 n，转化为通过均值点，得到 $\bar{y} = \beta_1 \bar{x} + \beta_0$

同时通过一阶条件还可以得到：

$$\frac{\partial SSR}{\partial \beta_1} = -2 \sum (y - \beta_1 x - \beta_0) x = 0$$

将 $\beta_0 = \bar{y} - \beta_1 \bar{x}$ 代入消元。

最终解得：

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}.$$

一些需要知道的**数学性质**：

定理 2.2 (OLS 公式重要的数理性质)
因为 $\sum_{i=1}^{n}(x_i-\bar{x})=0$, 所以可以得到:

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

$$\sum_{i=1}^{n}(x_i-\bar{x})x_i=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

♡

证明 关于离合差公式的证明

$$\begin{aligned}\sum_{i=1}^{n}x_i(y_i-\bar{y})&=\sum_{i=1}^{n}\left[(x_i-\bar{x})+\bar{x}\right](y_i-\bar{y})\\&=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+\bar{x}\sum_{i=1}^{n}(y_i-\bar{y})\\&=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}),\end{aligned}$$

因为 $\sum\limits_{i=1}^{n}(y_i-\bar{y})=0.$
同理,

$$\begin{aligned}\sum_{i=1}^{n}x_i(x_i-\bar{x})&=\sum_{i=1}^{n}\left[(x_i-\bar{x})+\bar{x}\right](x_i-\bar{x})\\&=\sum_{i=1}^{n}(x_i-\bar{x})^2+\bar{x}\sum_{i=1}^{n}(x_i-\bar{x})\\&=\sum_{i=1}^{n}(x_i-\bar{x})^2,\end{aligned}$$

因为 $\sum\limits_{i=1}^{n}(x_i-\bar{x})=0$。

笔记 上述的数学性质都依赖 $\sum_{i=1}^{n}(x_i-\bar{x})=0$,所以如果只有 $\sum xy$ 时不能使用。
借助以上式子,最小二乘的系数估计就可以变为:

$$\hat{\beta}_1=\frac{\sum(x-\bar{x})(y-\bar{y})}{(x-\bar{x})^2}=\frac{s_{xy}}{s_x^2}=\frac{\mathrm{Cov}(x,y)}{\mathrm{Var}(x)}$$

而前文提到的相关性式子为

$$r_{xy}=\frac{s_{xy}}{s_xs_y}=\frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}}=Cor(x,y)$$

笔记 相关性和因果性 一般认为,相关性和因果性的差别是因果性能明确判断方向性。相关性只能表达 x 和 y 同向变化,而因果性能进一步分析出是 x 指向 y 导致的变化。
将相关分析和最小二乘的公式进一步对比转化,可以得到:
回归分析的差异就是在相关性分析(r_{xy} 对称关系)的基础上重视方向 $r_{xy}\cdot\frac{s_y}{s_x}$(x 对 y 的贡献)。
注[相关性和因果性] 结论2.3.2是有限制的。

$$\beta = \frac{s_{xy}}{s_x s_y} = r_{xy}\frac{s_y}{s_x}$$

此对比结论只在一元线性回归中成立！不过多元线性回归可以在这个形式上变形拓展。

命题 2.1 (补充更深的理解)

补充，均方误思想只是一种统计思想，其形式本质还可以结合条件期望放到测度论中理解。

$$E\left[(X - E[X|\mathcal{G}])^2\right] = \min\{E(X-Y)^2 : Y \in \mathcal{G}, E(Y^2) < \infty\}$$

在 OLS 中，其表现为：

$$E\left[(Y - E[Y|X])^2\right] = \min\{u^2\}$$

推荐参考：测度论笔记（六）：独立性、条件期望、一致可积性

正是这个思想导致了 y 的方差就是 u 的方差。

此时先从直觉上理解下，y 的拟合就是 βx, 因此，残差 u 其实就是实际值 y 减去拟合值 $\beta x = \hat{y}$, 也可以看作 y 的均值。

问题 2.2 如果没有截距项（一元的情况）

这里是默认设置了截距 β_0 的求解情况，如果没有截距的情况呢？实际上，没有截距的情况就是《计量经济学导论（第六版）》46 页的情况。最终结果表示结果并不一样。

最终结果是

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

注 同时请注意，无截距的公式 $\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$ 不能转化成有截距的公式的 $\hat{\beta}_1 = \frac{\sum (x_i-\bar{x})y_i}{\sum (x_i-\bar{x})^2}$。

对应的数学性质2.3.2只能在 xy 当中有一个已经是 $\sum(x-\bar{x})$ 的情况下使用。

实际上，如果模型实际上是有截距的情况下，无截距回归含义是有遗漏变量 β_0。此时估计为

$$\hat{\beta}_1 = \beta_1 + E\left[\frac{\sum xu}{\sum x^2}\right] = \beta_1 + \beta_0\frac{\sum x_1}{\sum x_1^2}$$

截距的经济意义就是全部解释变量为 0 时因变量的基础值。例如婴儿一出生就有身高。从式子也可以看出，当截距为 0 时，无截距模型和有截距模型结果相同。如果没有截距，方程必然经过原点。

这里也能理解——常数项带有去中心化的意义。在一阶条件中，对截距常数项 β_0 求导，得到的意义是拟合方程经过均值。

2.3.3 拟合优度

关于拟合优度，实际上我们需要明白的是怎么设计出一套模型准确性的评价标准。

多元线性回归的总变差平方和依然满足如下关系：

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

- **总变差 SST：** $\sum(y_i - \bar{y})^2$
- **可解释部分 SSE：** $\sum(\hat{y}_i - \bar{y})^2$
- **未解释部分 SSR：** $\sum(y_i - \hat{y}_i)_1^2$

- 可决系数 $R^2 == \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$
- 修正可决系数 $\overline{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\sum \widehat{u}_i^2/(n-k-1)}{\sum (y_i-\overline{y})^2/(n-1)}$
- 可决系数和修正可决系数的关系：$\overline{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k-1}$

调整 R^2 小于等于 R^2

证明

$$\begin{aligned}
\overline{R}^2 &= 1 - \frac{\mathrm{SSR}/(n-k-1)}{\mathrm{SST}/(n-1)} \\
&= 1 - \frac{n-1}{n-k-1} \cdot \frac{\mathrm{SSR}}{\mathrm{SST}}, \\
R^2 - \overline{R}^2 &= \left(\frac{n-1}{n-k-1} - 1\right)\frac{\mathrm{SSR}}{\mathrm{SST}} \\
&= \frac{k}{n-k-1} \cdot \frac{\mathrm{SSR}}{\mathrm{SST}} \quad \geq 0,
\end{aligned}$$

$$\sum_i^n (y_i - \bar{y})^2 = \sum_i^n (\hat{y}_i - \bar{y})^2 + \sum_i^n (y_i - \hat{y})^2$$

$$SST = SSR + SSE$$

$$MSE = \frac{1}{n}(y_i - \bar{y})^2$$

$$RMSE = \sqrt{MSE}$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = r_{xy}^2 = 1 - \frac{MSE}{VAR(y)}$$

2.3.4 矩估计

定义 2.2 (矩估计)

矩估计只是一种表现方法，是基于样本矩去拟合。除了矩估计之外，还有很多估计方法，例如极大似然 (基于分布去拟合)。

其核心思想就是对样本总体期望进行约束，

$$\mathbb{E}[g(Z_i, \theta_0)] = 0$$

Z_i: 观察到的数据向量 (可能包括 y_i, x_i, z_i) $\cdot$ θ_0 : 未知参数向量

$g(\cdot)$: 称为矩函数 (moment function), 是根据模型和假定构造的

在计量中自然就表现为未观察到的误差项和解释变量之间的期望约束关系。♣

在 ols 中，对应的**矩条件**是 $Cov(x,u) = 0$, 或者写成 $E(u|x) = 0$

$$Cov(x,u) = Cov(x, y - \beta x) = Cov(x,y) - \beta Var(x)$$

同样可以解得：

$$\hat{\beta} = \frac{Cov(x,y)}{Var(x)} = \beta + \frac{Cov(y,x)}{Var(x)}$$

还可以写成矩阵形式

计量经济学导论附录 E

$$\hat{\beta} = (X'X)^{-1}X'Y$$

注[使用矩阵表示] 注意

- 要注意 $\hat{\beta}$ 为估计量，帽子不能丢
- 要注意行列式的表示，例如 X_i 是列向量，X_i' 为行向量，先后顺序需要保持行向量和列向量，放前面就是行向量，放后面就是列向量。
- 矩阵和方程等价：$X'X = \sum x_i x_i$

2.3.4.1 假设

命题 2.2 (几个假设)

1. 线性于参数: 计量线性的概念相对于参数存在。
2. 随机抽样: 随机抽样和非随机抽样（简单、系统、分层、整群、多阶段）。本质是样本独立同分布。
3. 无完全多重共线性: 允许自变量之间存在相关关系, 但不允许完全相关。
4. 零条件期望: 有误差项自动满足。残差没有经济含义，只代表真实值和测量值的差异，误差有经济含义。
5. 条件同方差性、误差项之间无自相关
6. 正态性

笔记 计量假设的理解

线性于参数。为了求 ols 的解析解。如果不是线性参数就使用二次形式，但是求解更为复杂。同时，线性参数有斜率的含义。

随机抽样意味着独立同分布。这使得最终可以使用大数定理和中心极限。

（独立性）误差项之间无自相关：各自的误差项（未观察到的变量）不相关（随机抽样时满足）。时间序列时容易不成立，趋势、周期性（例如周期）。违背了独立性依旧满足无偏，但不再是最有效的。

0 条件均值本质假定的是条件期望为 0——$E(u|x) = 0$，其实暗含了一种外生性假设：残差和解释变量不相关（但外生性存在强弱之分）。

（同分布）条件同方差。同方差假设的理解是条件同方差。从一个总体抽取不同的值，方差是相同的，但实际上条件方差可能不同。例如抽取一个学校的成绩，不同年级的成绩可能方差不同。

其中高斯马尔科夫假设只需要线性于参数、随机抽样、解释变量有波动、零条件均值、同方差性，ols 就是线性最优无偏估计（BLUE），加上正态性是为了进行精准的推断统计（构造统计量）。

假设并不影响计量参数值的估计，但影响参数值是否是有效的。即便不满足假设，也能求解 ols，但是如果满足假设，求解结果是非常有效的。

2.3.4.2 外生性和内生性

区分弱外生性、强外生性。

弱外生性：$E[u_t \mid X] = 0$。

强外生性：

误差和 x 的各种函数形式也不相关。

。

笔记 理解 0 条件均值三者的关系是什么？

$$(1)E[u \mid X]=0, \quad (2)E[Xu]=0, \quad (3)\operatorname{Cov}(X,u)=0$$

$$E[u \mid X]=0 \quad \Rightarrow \quad E[Xu]=0 \quad \Leftrightarrow \quad \operatorname{Cov}(X,u)=0 \text{ (if } E[u]=0)$$

如果满足 0 条件期望 $E[\varepsilon_t \mid X_t]=0$, 便可以使用迭代期望法则：

$$E[X\varepsilon]=E[E[X\varepsilon \mid X]]=E[XE[\varepsilon \mid X]]=E[X\cdot 0]=0$$

反过来，$E[\varepsilon_t * X_t]=0$ 只能保证线性不相关，而不能保证非线性不相关。

也可以从下面转化理解：

$$\operatorname{cov}(x,u)=E[(x-E[x])u]=E[(x-E[x])E[u \mid x]]=0$$

也就是说 $E[u_t \mid X]=0 \longrightarrow cov(x,u)=0$, 但是反之不成立。

$$\operatorname{Cov}(X,u)=E[(X-E[X])(u-E[u])]=0$$

$$\operatorname{Cov}(X,u)=E[Xu]-E[X]E[u]$$

第三章 多变量 OLS

多元回归的优势：

y 本质是一种条件期望，x 的变动影响的是 y 的条件期望。$E(Y|X) = f(X)$

- **可以在回归时使更多的因素保持不变**。控制更多的影响因素可以在回归分析中保持这些因素不变，而只关注核心解释变量的变动对被解释变量的影响。
- **更有可能满足零条件均值假定**。假如能把 u 中与 x 相关的因素分离出来, 直接作为解释变量, 那么, 零条件均值假定便可满足。
- **可以提高预测精度**。增加解释变量总可以提高, 由此可以更好地解释总体变异。可以灵活设定总体回归函数。比如, 在模型中引入解释变量的二次项、交叉项等。

FWL 本质是**分步回归**。现实世界充满了不可表示的变量，此时 FWL 在证明中就很有用了。

3.1 Frisch-Waugh 定理

Frisch-Waugh 定理告诉我们的是为何能多变量回归能“保持其他变量不变”。

3.1.1 方程形式

注 保持其他变量不变的一种**代数理解**：

$$\widehat{\beta}_j = \frac{\sum \widehat{\gamma}_{ij} y_i}{\sum \widehat{\gamma}_{ij}^2}(j = 1, 2, \cdots, k)$$

$\hat{\gamma}_{ij}(j = 1, 2, \cdots, k)$表示$x_j$ 对其它解释变量做 OLS 估计后得到的残差。

因此,$\hat{\gamma}_{ij}(j = 1, 2, \cdots, k)$ 表示 x_j 排除了或净化掉其它解释变量的影响后的部分。于是,$\widehat{\beta}_j$ 便度量了在排除其它解释变量后 y 和 x_i 之间的关系。

证明 [方程形式的 fwl]

对于多元回归，先得到一阶条件：

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_k) = 0 \tag{3.1}$$

$$-2\sum_{i=1}^{n} x_{i1}\,\underbrace{(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \sum_{j=2}^{k}\hat{\beta}_j x_{ij})}_{\hat{u}_i} = 0 \tag{3.2}$$

接下来将 x_{i1} 对其他控制变量回归——$x_{i1} = \hat{x}_{i1} + \hat{r}_{i1}$。

同时回归满足正交性质：$\hat{x}_{i1}\hat{r}_{i1} = 0$。

代入一阶条件(3.2)：

$$\sum_{i=1}^{n}(\hat{x}_{i1} + \hat{r}_{i1})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_k) = 0$$

由于误差项外生性性质 $\sum_{i=1}^{n}\hat{x}_{i1}\hat{u}_i = 0$

式子存在两个残差 × 解释变量。

一个是 $\hat{x}_{i1}$ 和 $(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_k)$;

一个是 $\hat{r}_{i1}$ 和一系列 x 解释变量。全部交乘之和为 0。

最终化简得到：

$$\sum_{i=1}^{n} \hat{r}_{i1}(y_i - \hat{\beta}_1 \hat{r}_{i1}) = 0$$

多元回归中, 控制变量和解释变量在回归式统计地位上是等价的, 但我们对解释变量更干兴趣, 加入控制变量的目的是为研究解释变量的边际作用。

部分回归的基本思想是, 当引入控制变量后, 若想探究解释变量 x 与被解释变量 y 的相关系数, 那么就先剔除 (partial out) 控制变量对 y 的影响和剔除控制变量对 x1 的影响, 之后再让剩余部分的 y 对剩余部分的 x1 做回归。

代码例子

推荐更简洁的参考资料 1和参考资料 2

Listing 3.1: fwl 验证代码

```
*fwl形式1: x1对其他解释变量回归, 生成残差e; y对e回归
sysuse auto, clear
reg mpg weight length foreign
reg weight length foreign
predict gamma_w, resid
reg mpg gamma_w
*fwl形式1: x1对其他解释变量回归, 生成残差e; y对e回归
sysuse auto, clear
reg mpg weight length foreign
reg weight length foreign
predict gamma_w, resid
reg mpg length foreign
predict gamma_y, resid
reg gamma_y gamma_w
```

问题 3.1 如果没有截距

在问题2.2中，有截距和无截距，ols 的解形式不同，差别在于是否去中心化 $x - \bar{x}$。

多元回归时，有截距和无截距形式一样 ($\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$)，只是代入 y 时有无截距的区别。原因就是在于多元回归控制了其他变量不变。残差在排除其他 x 的影响过程中，也排除掉了常数项，因此形式上没有去中心化这个过程，因为残差 $\hat{r}_{i1}$ 本身就类似于 $x - \bar{x}$。

3.1.2 矩阵形式

证明 [FWL]

已知 OLS 一阶条件计算 $\hat{\beta} = (X'X)^{-1}X'y$

定义原始回归 $y = \beta_1 x_1 + \beta_2 + x_2 + u$, 得到原始式的估计为 $\hat{\beta}_1$ 和 $\hat{\beta}_2$。

使用 X_2 对 X_1 做回归 $X_2 = \ddot{\beta}_2 X_1 + \ddot{X}_2$,

得到残差 $\ddot{X}_2$,

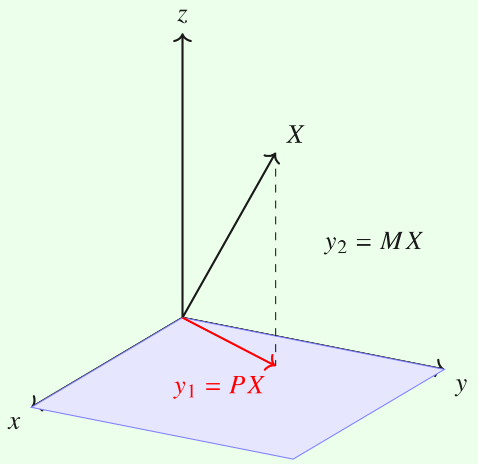
$$\ddot{\mathbf{X}}_2 = \left[\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right]\mathbf{X}_2 = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}_2 = \mathbf{M}_1\mathbf{X}_2$$

并在这个过程中得到投影矩阵 $P \equiv X(X'X)^{-1}X'$ 和消失矩阵 $M \equiv I - P$。

注意变形：对于 $y = \beta x + u$ 有 $\hat{y} = \hat{\beta}X = Py$ 和 $u = y - \hat{y} = My$

注 其中 P 和 M 均代表幂等矩阵, 特点是 $AA' = A$

其中 P 是投影矩阵,P 左乘任何向量会得到平面上的投影。
M 为消灭矩阵,M 左乘任何向量会得到本身减去投影后的剩余向量。



接下来研究原始回归式:

$$Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + u$$

首先有一阶条件公式:$X'X\hat{\beta} = X'Y$
然后在回归式子中有 X_1X_2 两个变量, 因此转化为矩阵。

$$\begin{aligned} X'X = & \quad (X_1 \quad X_2)'(X_1 \quad X_2) \\ = & \begin{pmatrix} X_1{}' \\ X_2{}' \end{pmatrix}(X_1 \quad X_2) \\ = & \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \end{aligned}$$

同理, 接下来是系数 β

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

最后是 $X'Y$

$$\begin{aligned} X'Y &= (X_1X_2)'Y \\ &= \begin{pmatrix} X_1'Y \\ X_2'Y \end{pmatrix} \end{aligned}$$

最终得到:

$$\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'Y \\ X_2'Y \end{pmatrix}$$

也就是两个等式

$$X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 = X_1'Y \tag{3.3}$$
$$X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 = X_2'Y \tag{3.4}$$

解第方程3.3得到

$$\hat{\beta}_1 = (X_1'X_1)^{-1}(X_1'Y - X_1'X_2\hat{\beta}_2) = (X_1'X_1)^{-1}X_1'(Y - X_2\hat{\beta}_2) \tag{3.5}$$

将方程3.5代入方程3.4, 得到

$$X_2'X_1(X_1'X_1)^{-'}X_1'(Y - X_2\widehat{\beta}_2) + X_2'X_2\widehat{\beta}_2' = X_2'Y \tag{3.6}$$

对于方程3.6, 发现投影矩阵 $P_1 = X_1(X_1'X_1)^{-'}X_1'$。
于是式子转化为:

$$\begin{aligned} X_2'X_1(X_1'X_1)^{-'}X_1'(Y - X_2\widehat{\beta}_2) + X_2'X_2\widehat{\beta}_2' &= X_2'Y \\ X_2'P_1Y + X_2'P_1X_2\hat{\beta}_2 + X_2'X_2\hat{\beta}_2 &= X_2'Y \\ X_2'(I - P_i)X_2\hat{\beta}_2 &= X_2'(I - P_1)Y \\ X_2'M_1X_2\hat{\beta}_2 &= X_2'M_1Y \end{aligned}$$

也就得到

$$\hat{\beta}_2 = (X_2'M_1X_2)^{-1}X_2'M_1Y$$

这个形式与一阶条件非常相似——$\hat{\beta} = (X'X)^{-1}X'Y$
通过幂等矩阵的特性 $(M'M = M = M')$ 继续补全:

$$\begin{aligned} \hat{\beta}_2 &= (X_2'M_1X_2)^{-1}X_2'M_1Y \\ &= (X_2'M_1M_1X_2)^{-1}X_2'M_1M_1Y \\ &= [(M_1X_2)(M_1X_2)']^{-1}(M_1X_2)'(M_1Y) \end{aligned}$$

含义为:
先让 x_2 对 x_1 回归, 保留残差 $\ddot{x}_2$;
再用 y 对 x_1 回归, 保留残差 $\ddot{y}$;
再用 $\ddot{y}$ 对 $\ddot{x}_2$ 做回归, 会得到 $\hat{\beta}_2$;
如果保留用 $\ddot{y}$ 对 $\ddot{x}_2$ 做回归的残差 $\ddot{u}$, 还发现等于原始式子残差。
证明如下:

$$\ddot{y} = M_1y = M_1(\hat{\beta}_1X_1 + \hat{\beta}_2X_2 + \hat{u})$$

基于消失矩阵的性质 $M_1\hat{\beta}_1X_1 = 0$
得到

$$\ddot{y} = M_1y = M_1X_2\hat{\beta}_2 + M_1\hat{u}$$

同时因为 $X_1'\hat{u} = 0$

$$M_1\hat{u} = (\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1')\hat{\boldsymbol{u}} = \hat{u}$$

原式子变为

$$M_1y = M_1X_2\hat{\beta}_2 + \hat{u}$$

引理 3.1

基于 FWL, 可以得到引理。
此时保留 $\ddot{y}$ 对 $\ddot{x}_2$ 做回归的差分 $\ddot{u}$, 会得到以下结论:
使用 X_2 对 X_1 做回归 $X_2 = \ddot{\beta}_2X_1 + \ddot{X}_2$,
得到残差 $\ddot{X}_2$,

$$\ddot{\mathbf{X}}_2 = \left[\mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right]\mathbf{X}_2 = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}_2 = \mathbf{M}_1\mathbf{X}_2$$

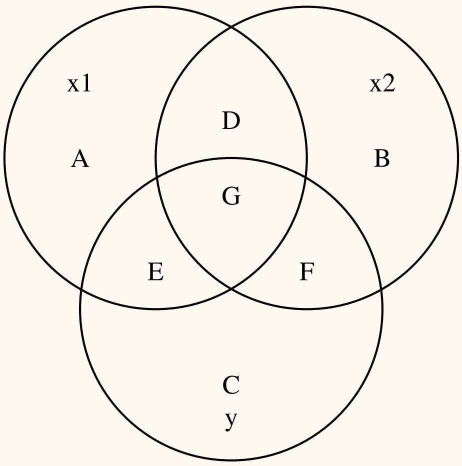
在上式中, 如果 $X_1'X_2 = 0$, 可得 $\ddot{X}_2 = M_1X_2 = X_2$
此时短回归式子 $y = \beta x_2 + u$ 中,

$$\tilde{\beta} = (X_2'X_2)^{-1}X_2'Y$$

而在回归式子 $y = \beta\ddot{x}_2 + u$ 中 ($\ddot{x}_2$ 为 X_2 对 X_1 回归的残差),

$$\ddot{\beta} = (\ddot{X}_2{}'\ddot{X}_2)^{-1}\ddot{X}_2{}'Y$$

此时长回归和短回归系数相等 $\ddot{\beta} = \tilde{\beta} = \hat{\beta}$
因此添加控制变量如果不影响显著数值, 原因就是控制变量和解释变量之间不相关。



如图, 就像这个集合,x1 和 x2 各自对 y 有影响, 但我们想研究的是分别是 E 和 F, 因此应当排除 G。
当 G=0 时, 长回归和短回归系数一样, 当 G≠0 时, 控制变量的加入会影响显著水平。

证明 可以通过 FWL 定理进行证明。

$$\hat{\beta}_K = \beta_K + [X_K'M_KX_K]^{-1}X_K'M_KU$$

接下来看其条件方差

$$Var(\beta_K|X) = \frac{\delta^2}{X_K'M_KX_K}$$

同时有以下变形

$$X_K' M_K X_K = [M_K X_K]'[M_K X_K] = SSE_K$$

最终变形得到

$$X_K' M_K X_K = \frac{\sigma^2}{(n-1)s_K^2(1-R_K^2)}$$

其中，$s_K^2 = (n-1)^{-1}\sum_{i=1}^n (x_{iK} - \bar{x}_K)^2$ 为 x_K 的样本方差。

多重共线性问题是当 $R_K^2 \to 1$ 时，$\mathrm{var}(\beta_K \mid X) \to \infty$：如果第 K 个解释变量几乎与其他解释变量共线（即几乎是其他解释变量的线性组合），那么 $\mathrm{var}(\beta_K \mid X)$ 将会变得非常大，因此 $\mathrm{SE}(\beta_K)$ 将会变得很大。另一方面，R_K^2 只是决定 $\mathrm{var}(\beta_K)$ 的其中一个因素，其他影响 $\mathrm{var}(\beta_K)$ 的因素还有样本方差 s_K^2 及样本容量 n，原则上都可以补偿多重共线性所带来的问题。总之，高度共线的变量会使 SE 膨胀，但是，这不会导致偏误或不一致估计。其实样本越大，自变量的变化越大，估计越精确。

解决办法：**此时考虑线性组合两个相关变量。**

3.2 期望

3.2.1 方程形式

方程形式以多元线性回归为例子：

$$\widehat{\beta}_j = \frac{\sum_i \widehat{\gamma}_{ij} y_i}{\sum_i \widehat{\gamma}_{ij}^2} = \frac{\sum_i \widehat{\gamma}_{ij}(\sum_l \beta_l x_{il} + u_i)}{\sum_i \widehat{\gamma}_{ij}^2}$$

因为 $\sum_i \widehat{\gamma}_{ij} x_{il} = 0 (l \neq j)$。

$$\widehat{\beta}_j = \frac{\sum_i \widehat{\gamma}_{ij} y_i}{\sum_i \widehat{\gamma}_{ij}^2} = \frac{\sum_i \widehat{\gamma}_{ij}(\sum_l \beta_l x_{il} + u_i)}{\sum_i \widehat{\gamma}_{ij}^2} = \frac{\sum_i \widehat{\gamma}_{ij}(\beta_j x_{ij} + u_i)}{\sum_i \widehat{\gamma}_{ij}^2} = \beta_j \frac{\sum_i \widehat{\gamma}_{ij} x_{ij}}{\sum_i \widehat{\gamma}_{ij}^2} + \frac{\sum_i \widehat{\gamma}_{ij} u_i}{\sum_i \widehat{\gamma}_{ij}^2}$$

又因为 $x_{ij} = \delta_0 + \sum_{l \neq j} \delta_1 x_{il} + \ldots + \widehat{\gamma}_{ij}$, 也就是

$$x_{ij}\widehat{\gamma}_{ij} = [\delta_0 + \sum_{l \neq j} \delta_1 x_{il} + \ldots + \widehat{\gamma}_{ij}]\widehat{\gamma}_{ij} = \widehat{\gamma}_{ij}^2$$

最终，

$$\widehat{\beta}_j = \beta_j \frac{\sum_i \widehat{\gamma}_{ij}^2}{\sum_i \widehat{\gamma}_{ij}^2} + \frac{\sum_i \widehat{\gamma}_{ij} u_i}{\sum_i \widehat{\gamma}_{ij}^2} = \beta_j + \frac{\sum_i \widehat{\gamma}_{ij} u_i}{\sum_i \widehat{\gamma}_{ij}^2}$$

$$E[\widehat{\beta}_j \mid X] = \beta_j + E\left[\frac{\sum_i \widehat{\gamma}_{ij} u_i}{\sum_i \widehat{\gamma}_{ij}^2} | X\right] = \beta_j + \frac{\sum_i \widehat{\gamma}_{ij} E[u_i \mid X]}{\sum_i \widehat{\gamma}_{ij}^2} = \beta_j$$

3.2.2 矩方法

矩方法就是使用样本均值替换估计量进行估计，于是得到:

此处得带求解的 **OLS 假设的 0 条件均值**：$E(x'u) = 0$

注 残差平方可以写为矩阵形式。

$$SSR = \sum_{t=1}^{n} = \hat{u}'\hat{u} = (Y - X\hat{\beta})(Y - X\hat{\beta})'$$

定理 3.1 (OLS 无偏性)

$$
\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}
\end{aligned}
$$

接下来求基于 X 的条件期望, 此时有假定 $E(u|x) = 0$

$$
\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\mathbf{u} \mid \mathbf{X}) \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \beta
\end{aligned}
$$

注 无关变量的估计系数为 0。如果加入了无关变量, 也就是 x 和 y 无关的变量, 不会影响各自估计的无偏性。

注[遗漏变量]

接下来研究**遗漏变量**的情形。

原始回归为:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

而我们遗漏了变量 x_2, 只对 x_1 进行了简单回归, 使用如下形式表示:

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

为了进一步研究偏差了多少, 我们使用 x_2 对 x_1 回归, 得到斜率 $\tilde{\delta}_1$

也就是

$$x_2 = \bar{\delta}_1 x_1 + u$$

如果 x_2 是 y 的无关变量, 则 $\beta_2 = 0$; 如果 x_2 与 x_1 不相关, 则 $\tilde{\delta}_1 = 0$。

接下来研究无偏估计:

$$
\begin{aligned}
E(\tilde{\beta}_1 &= E(\hat{\beta}_1) + \beta_2\tilde{\delta}_1) = E(\hat{\beta}_1) + E(\hat{\beta}_1) + E(\hat{\beta}_2)\tilde{\delta}_1 \\
&= \beta_1 + \beta_2\tilde{\delta}_1
\end{aligned}
$$

$\beta_2\tilde{\delta}_1$ 就是遗漏变量偏误。

3.3 方差

3.3.1 方程形式

以多元函数为例子, 基于 fwl 定理3.1, 此时 β 求解为

$$\hat{\beta}_j = \frac{\sum_i \gamma_{ij}\,\tilde{y}_i}{\sum_i \gamma_{ij}^2} = \beta_j + \frac{\sum_i \gamma_{ij}\,\tilde{u}_i}{\sum_i \gamma_{ij}^2}$$

接下来求解对应方差:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_j \mid X) &= \mathrm{Var}\left(\frac{\sum_i \gamma_{ij}\hat{u}_i}{\sum_i \gamma_{ij}^2} \,\middle|\, X\right) \\ &= \frac{1}{\left(\sum_i \gamma_{ij}^2\right)^2}\sum_i \gamma_{ij}^2\ \mathrm{Var}(\hat{u}_i \mid X) \ = \frac{\hat{\sigma}^2}{\sum_i \gamma_{ij}^2}.\end{aligned}$$

笔记 [展开说明] 中间方差的展开利用了 $\mathrm{Var}(a+b) = \mathrm{Var}(a) + \mathrm{Var}(b) + 2\mathrm{Cov}(a,b)$。

笔记

总体方差不是估计量，因为无法估计，σ 只能通过样本方差 $\hat{\sigma}$ 逼近。同理，总体样本的 $\mathrm{Var}(\beta_j \mid X)$ 也不是估计量。我们汇报的是标准误 $\mathrm{Var}(\hat{\beta}_j \mid X)$, 也就是 $se(\hat{\beta})$

笔记

实际上样本和总体公式的系数估计相同，但是在计算方差时很重要。

估计量表示：

$$\widehat{\mathrm{Var}}(\hat{\beta}_j \mid X) = \frac{\hat{\sigma}^2}{\sum_i \gamma_{ij}^2}, \qquad \hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-k}.$$

对于残差 γ_{ij}, 是 x_j 对剩余解释变量的回归的残差，有：

$$\begin{aligned}SST_T &= \sum_i (x_{ij} - \bar{x}_j)^2 \\ SSE_j &= \sum_i (x_{ij} - \hat{x}_{ij})^2 = \sum_i \gamma_{ij}^2 \\ SSR_j &= \sum_i (\hat{x}_{ij} - \bar{x}_j)^2 \\ SST_j &= SSR_j + SSE_j \\ R_j^2 &= \frac{SSR_j}{SST_j}\end{aligned}$$

于是得到

$$\sum_i \gamma_{ij}^2 = SSE_j = SST_j - SSR_j = (1 - R_j^2)SST_j$$

也就是：

$$Var(\hat{\beta}_j|x) = \frac{\sigma^2}{(1-R_j^2)SST_j}, j = 1, \cdots, k$$

注共线性的理解

注意，这里的 R^2 和 SST 都是 x_j 对其他解释变量 x_i 回归的一系列结果。

误差方差 $\hat{\sigma}$ 越大，ols 估计量方差越大。

这个式子可以看出，当 x 存在共线越强时时，R^2 增大，SST(总样本变异) 变小（一个解释变量几乎能解释另外一个解释变量）。因此方差会变得很大。这也说明了解释变量要选取多维度的信息。

3.3.2 矩估计

假定 5. 同方差性和不存在序列自相关

是为了获取 $\hat{\beta}$ 最简单的协方差矩阵。

(i)$\mathrm{Var}(u_i|\mathbf{X}) = \sigma^2; t = 1, 2, \cdots; n$。

同方差假定,u_t 的方差不依赖于 X 中的任意一个元素, 且不同观测 t 的方差都相等。

异方差示意图,

由于 $var(u|x) = var(y|x)$, 因此只要 $var(u|x)$ 是 x 的函数, 就会出现**异方差**。

证明 证明 $var(u|x) = var(y|x)$

首先

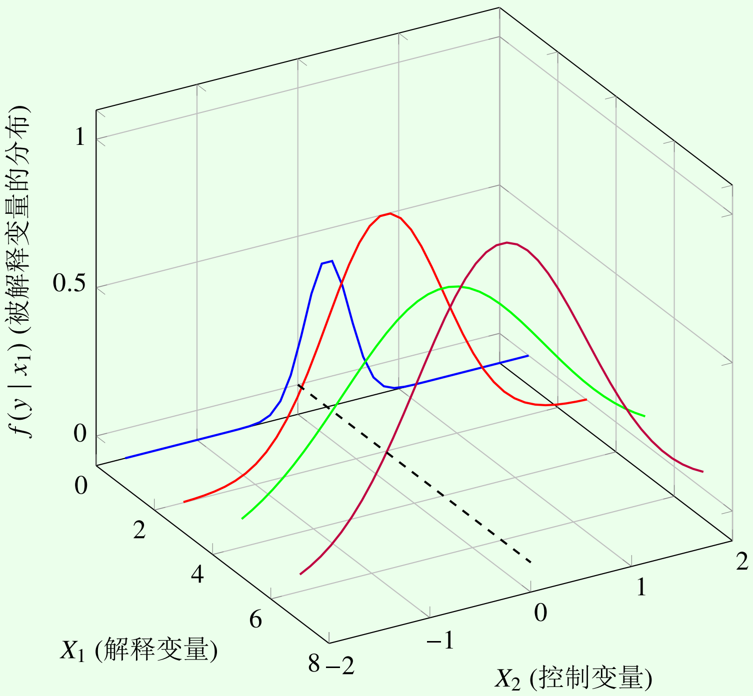
$$E(Y|X) = E(\beta_0 + \beta_1 X + u|X) = \beta_0 + \beta_1 X$$

且有零条件均值假设 $E(u|X) = 0$

接下来展开 $Var(Y|X) = E((Y - E(Y|X))|X)$

$$\begin{aligned} Var(Y|X) &= E((Y - E(Y|X))|X) = E((Y - \beta_0 - \beta_1 X)^2|X) \\ &= E(u^2|X) = E(u - 0)^2|X) = E(u - E(u|X))^2|X) \\ &= Var(u|X) \end{aligned}$$

异方差示意图:



(ii)$\mathrm{Cov}(u_i, u_s|\mathbf{X}) = 0$ 对任意的 $t \neq s$ 都成立。

不同观测点误差不能相关。在时间序列中, 这点表现为不同时期的误差不相关。

以矩阵形式, 我们可以将这两个假定写成 $Var(u|X) = \sigma^2 I_t$。其中,I_n 是 n×n 单位矩阵。

以上假设成立, 则 u 具有数量方差-协方差矩阵。现在就可以进一步推导 OLS 估计量是方差-协方差矩阵论。

有 $var(u \mid x) = E(u^2 \mid x) - [E(u \mid x)]^2$ 和 $E(u \mid x) = 0$,

于是得到：$E(u^2 \mid x) = var(u \mid x) = \sigma^2$, 也就说明 $E[E(u^2 \mid x)] = E(u^2) = \sigma^2$。也就说明 σ^2 是 u 的无条件方差。

于是接下来我们研究 β 的方差

定理 3.2 (OLS 估计量的方差-协方差)

基于线性于参数、不存在完全共线、零条件均值、**OLS** 无偏性、同方差和不存在序列相关的假设，有以下结论：

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

证明

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\end{aligned}$$

基于这个式子估计最小二乘估计量的方差。

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= \mathrm{E}((\hat{\beta} - E(\hat{\beta}_0|X))(\hat{\beta} - E(\hat{\beta}_0|X))' \mid \mathbf{X}) \\ &= \mathrm{E}((\hat{\beta} - \beta))(\hat{\beta} - \beta))' \mid \mathbf{X}) \\ &= \mathrm{E}(((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})' \mid X) \\ &= (X'X)^{-1}X'E\left[uu'|X\right]X\left(X'X\right)^{-1} \\ &= \mathrm{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \mid \mathbf{X}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathrm{Var}(\mathbf{u} \mid \mathbf{X})]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

再基于同方差假定 $Var(u|X) = \sigma^2 I_n$

对式子进行进一步的简化即可证明

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

定义 3.1

假定 1 到假定 5 联合起来就是 (横截面的) 高斯马尔科夫假定

- 假定 1：线性于参数
- 假定 2：随机抽样，含义是独立同分布 iid
- 假定 3:0 条件均值
- 假定 4：不完全多重共线性
- 假定 5：同方差

♣

3.4 高斯马尔科夫定理

3.4.1 方程形式

《计量经济学导论》第三章附录最后一个证明（92 页）。

假设存在一个参数 $\tilde{\beta}$

证明

$$\begin{aligned}\tilde{\beta}_1 &= \beta_0\sum_{i=1}^{n} w_{i1} + \beta_1\sum_{i=1}^{n} w_{i1}x_{i1} + \beta_2\sum_{i=1}^{n} w_{i1}x_{i2} + \cdots + \beta_k\sum_{i=1}^{n} w_{i1}x_k + \sum_{i=1}^{n} w_{i1}u_i\beta_1\\ &= \beta_0\sum_{i=1}^{n} w_{i1} + \beta_1\sum_{i=1}^{n} w_{i1}x_{i1} + \beta_2\sum_{i=1}^{n} w_{i1}x_{i2} + \cdots + \beta_k\sum_{i=1}^{n} w_{i1}x_k + \sum_{i=1}^{n} w_{i1}u_i\end{aligned}$$

满足 2-4 假定的情况下，

$$\begin{aligned}\mathbf{E}(\tilde{\beta}_1|\mathbf{X}) = \beta_0\sum_{i=1}^{n} w_{i1} + \beta_1\sum_{i=1}^{n} w_{i1}X_{i1} + \beta_2\sum_{i=1}^{n} w_{i1}x_{i2} + \cdots\\ + \beta_k\sum_{i=1}^{n} w_{i1}x_k + \sum_{i=1}^{n} w_{i1}\,\mathrm{E}(u_i \mid \mathbf{X})\\ = \beta_0\sum_{i=1}^{n} w_{i1} + \beta_1\sum_{i=1}^{n} w_{i1}x_{i1} + \beta_2\sum_{i=1}^{n} w_{i1}x_{i2} + \cdots + \beta_k\sum_{i=1}^{n} w_{i1}x_{ik}\end{aligned}$$

对于无偏性，有以下条件；

$$\sum_{i=1}^{n} w_{i1} = 0, \sum_{i=1}^{n} w_{i1}x_{i1} = 1, \sum_{i=1}^{n} w_{i1}x_{ij} = 0, j = 2, \cdots, k$$

接下来分析方差：

基于 fwl3.1定理，使用 $\hat{\gamma_{i1}}$ 表示 x_{i1} 对剩余解释变量的回归的残差。

同时因为 $\sum_{i=1}^{n} w_{i1}\hat{x}_{i1} = 0$。

展开 $\sum_{i=1}^{n} w_{i1}x_{i1} = 1$ 再简化得到 $\sum_{i=1}^{n} w_{i1}\hat{r_{i1}} = 1$。

$$\mathrm{Var}(\tilde{\beta}_1 \mid \mathbf{X}) - \mathrm{Var}(\tilde{\beta}_1 \mid \mathrm{X}) = \sigma^2\sum_{i=1}^{n} w_{i1}^2 - \sigma^2/(\sum_{i=1}^{n} r_{i1}^2)$$

接下来利用 $\sum_{i=1}^{n} w_{i1}r_{i1} = 1$，平方，然后代入第二项的分子：

于是得到

$$\sum_{i=1}^{n} w_{i1}^2 - (\sum_{i=1}^{n} w_{i1}r_{i1})^2/(\sum_{i=1}^{n} r_{i1}^2) \tag{3.7}$$

证明高斯马尔科夫定理就是证明这个式子大于等于 0。

方法 1：利用柯西不等式 2.1

已知：

$$\sum_{i=1}^{n} w_{i1}r_{i1} = 1$$

利用柯西-施瓦茨不等式：

$$\Big(\sum_{i=1}^{n} w_{i1}r_{i1}\Big)^2 \le \Big(\sum_{i=1}^{n} w_{i1}^2\Big)\Big(\sum_{i=1}^{n} r_{i1}^2\Big)$$

代入 $\sum_i w_{i1}r_{i1} = 1$ 得：

$$1^2 = \Big(\sum_{i=1}^{n} w_{i1}r_{i1}\Big)^2 \le \Big(\sum_{i=1}^{n} w_{i1}^2\Big)\Big(\sum_{i=1}^{n} r_{i1}^2\Big)$$

两边同时除以 $\sum_i r_{i1}^2$：

$$\frac{1}{\sum_{i=1}^{n} r_{i1}^2} \le \sum_{i=1}^{n} w_{i1}^2$$

于是：

$$\sigma^2 \sum_{i=1}^{n} w_{i1}^2 - \frac{\sigma^2}{\sum_{i=1}^{n} r_{i1}^2} \geq 0$$

也就是：

$$\mathrm{Var}(\tilde{\beta}_1 \mid \mathbf{X}) - \mathrm{Var}(\hat{\beta}_1 \mid \mathbf{X}) \geq 0$$

放缩就可以得到对应的式子(3.7)，也证明了大于等于 0。

方法 2：回到回归式子的理解

$$\sum_{i=1}^{n} w_{i1}^2 - \frac{\left(\sum_{i=1}^{n} w_{i1} r_{i1}\right)^2}{\sum_{i=1}^{n} r_{i1}^2} = \sum_{i=1}^{n} w_{i1}^2 - \left(\frac{\sum_{i=1}^{n} w_{i1} r_{i1}}{\sum_{i=1}^{n} r_{i1}^2}\right)\left(\sum_{i=1}^{n} w_{i1} r_{i1}\right)$$

实际上可以得到一个系数 $\hat{\gamma}_1 = \frac{\sum_{i=1}^{n} w_{i1} r_{i1}}{\sum_{i=1}^{n} r_{i1}^2}$。

回归含义 $\hat{\gamma}_1$ 就是 w_{i1} 对 r_{i1} 的回归系数。

同时还得到了一个转化等式 $\sum_{i=1}^{n} w_{i1} r_{i1} = \hat{\gamma}_1 \sum_{i=1}^{n} r_{i1}^2$。

这个恒等式子还可以反复变化：

$-2\hat{\gamma}_1 \left(\sum_{i=1}^{n} w_{i1} r_{i1}\right) = -2\hat{\gamma}_1^2 \sum_{i=1}^{n} r_{i1}^2$

上面这个式子意味着以下转化：

$$-\hat{\gamma}_1^2 \sum r_{i1}^2 = -2\hat{\gamma}_1 \left(\sum w_{i1} r_{i1}\right) + \hat{\gamma}_1^2 \sum r_{i1}^2$$

接下来代入转化：

$$\begin{aligned}
\sum_{i=1}^{n} w_{i1}^2 - \left(\frac{\sum_{i=1}^{n} w_{i1} r_{i1}}{\sum_{i=1}^{n} r_{i1}^2}\right)\left(\sum_{i=1}^{n} w_{i1} r_{i1}\right) &= \sum_{i=1}^{n} w_{i1}^2 - \hat{\gamma}_1 \left(\hat{\gamma}_1 \sum_{i=1}^{n} r_{i1}^2\right) \\
&= \sum_{i=1}^{n} w_{i1}^2 - \hat{\gamma}_1^2 \sum_{i=1}^{n} r_{i1}^2 \\
&= \sum_{i=1}^{n} w_{i1}^2 - 2\hat{\gamma}_1 \left(\sum_{i=1}^{n} w_{i1} r_{i1}\right) + \hat{\gamma}_1^2 \sum_{i=1}^{n} r_{i1}^2 \\
&= \sum_{i=1}^{n} \left(w_{i1}^2 - 2 w_{i1} \hat{\gamma}_1 r_{i1} + \hat{\gamma}_1^2 r_{i1}^2\right) \\
&= \sum_{i=1}^{n} (w_{i1} - \hat{\gamma}_1 r_{i1})^2
\end{aligned}$$

3.4.2 矩估计

笔记 再次强调假设的理解

高斯马尔科夫的证明非常重要，在之后也会常常使用这样的半正定构造。

没有同方差，有线性无偏估计量中方差不再最小。

正态分布假设是为了构造统计量。小样本需要满足条件，大样本则有大数定理和中心极限定理作为支撑。

证明 β 的其他任何一个线性估计量都可以写成

$$\tilde{\beta} = A' y$$

为满足最小无偏条件估计，需要对 A 进行进一步限制。

$$\tilde{\beta} = \mathbf{A}'(\mathbf{X}\beta + \mathbf{u}) = (\mathbf{A}'\mathbf{X})\beta + \mathbf{A}'\mathbf{u}$$

接下来通过期望来研究无偏性

$$\begin{aligned}\mathbb{E}(\tilde{\beta}|\mathbf{X}) =& \mathbf{A}'\mathbf{X}\beta + \mathbf{E}(\mathbf{A}'\mathbf{u} \mid \mathbf{X}) \\ &= \mathbf{A}'\mathbf{X}\beta + A'\mathbf{E}(u|X) \text{因为 A 是 X 的一个函数} \\ &= \mathbf{A}'\mathbf{X}\beta, \text{ 因为} \mathbf{E}(\mathbf{u}|\mathbf{X}) = 0\end{aligned}$$

要满足无偏估计，则需要 $\mathbf{A}'\mathbf{X} = \mathbf{I}_{k+1}$(A 是一个 k*k 的矩阵)。接下来基于同方差假设，得到方差：

$$\operatorname{Var}(\bar{\beta}|\mathbf{X}) = \mathbf{A}'[\operatorname{Var}(\mathbf{u}|\mathbf{X})]\mathbf{A} = \sigma^2\mathbf{A}'\mathbf{A}$$

为比较有效性，将其和最小二乘的估计进行比较。

$$\begin{aligned}\operatorname{Var}(\bar{\beta} \mid \mathbf{X}) - \operatorname{Var}(\hat{\beta} \mid \mathbf{X}) &= \sigma^2\left[\mathbf{A}'\mathbf{A} - (\mathbf{X}'\mathbf{X})^{-1}\right] \\ &= \sigma^2\left[\mathbf{A}'\mathbf{A} - \mathbf{A}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\right], \text{ 因为} \mathbf{A}'\mathbf{X} = \mathbf{I}_{k+1} \\ &= \sigma^2\mathbf{A}'\left[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{A} \\ &\equiv \sigma^2\mathbf{A}'\mathbf{M}\mathbf{A} >= 0\end{aligned}$$

$\mathbf{A}'\mathbf{M}\mathbf{A}$ 是半正定矩阵，因此证明了最小二乘估计是最有效的。

定理 3.3

误差方差 σ^2 的无偏估计量可写成 $\sigma^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)$ 同样的假设条件下,$\hat{\sigma}$ 是 σ 的无偏估计。

证明

$$\hat{u} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = My = Mu$$

$My = Mu$ 是因为 $My = M(\beta X + u)$，因为 $MX = 0$(消失矩阵性质)，所以等号成立。

接下来使用幂等矩阵的性质，

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}\mathbf{u}$$

$\hat{\mathbf{u}}'\hat{\mathbf{u}}$ 最后会得到一个数，因此他本身等于迹 (对角线元素之和)。

$$\begin{aligned}\mathbf{E}(\mathbf{u}'\mathbf{M}\mathbf{u}|\mathbf{X}) &= \mathrm{E}[\operatorname{tr}(\mathbf{u}'\mathbf{M}\mathbf{u}) \mid \mathbf{X}] = \mathrm{E}[\operatorname{tr}(\mathbf{M}\mathbf{u}\mathbf{u}') \mid \mathbf{X}] \\ &= \operatorname{tr}[\mathrm{E}(\mathbf{M}\mathbf{u}\mathbf{u}') \mid \mathbf{X}] = \operatorname{tr}[\mathbf{M}\mathrm{E}(\mathbf{u}\mathbf{u}') \mid \mathbf{X}] \\ &= \operatorname{tr}(\mathbf{M}\sigma^2\mathbf{I}_n) = \sigma^2\operatorname{tr}(\mathbf{M}) = \sigma^2(n-k-1)\end{aligned}$$

最后一个等号成立原因

$$\operatorname{tr}(\mathbf{M}) = \operatorname{tr}(\mathbf{I}_n) - \operatorname{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] = n - \operatorname{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = n - \operatorname{tr}(\mathbf{I}_{k+1}) = n - (k+1)$$

完成总体误差方差的估计后，就可以估抽样方差了。

$\hat{\beta}_j$ 的标准差 $sd(\hat{\beta}_j)$ 的估计量如下所示：

$$se(\widehat{\beta}_j) = \frac{\widehat{\sigma}}{\sqrt{(1-R_j^2)SST_j}}, j = 1, 2, \cdots, k$$

$SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x})^2$，表示 x_j 总样本波动；R_j^2 是将 x_j 对所有其他自变量 (包含截距项) 回归得到的 R^2

$se(\widehat{\beta}_j)$ 被称为 $\widehat{\beta}_j$ 的**标准误** (standard error)。

3.5 系数含义

论文审稿，除了检验显著性，还需要审查其经济理论含义。例如只能为正的场合，系数居然为负。涉及一些变化率，只能在 0 到 1 之间而不能超过。或者估算的效率贡献为 95%，这显然是因为什么原因高估了。

此处注意区分单位、百分比、方向、缩放百分比。

取对数是 100% 比的含义

$$lny = \beta lnx + u$$

弹性：x 变化 1%,y 变化 β%

$$d(\ln y) = \beta d(\ln x) \Longrightarrow \frac{dy}{y} = \beta\frac{dx}{x} \Longrightarrow \Delta y\% \approx \beta\Delta x\%$$

$$y = \beta lnx + u$$

x 变化 1%,y 变化 $\frac{\beta}{100}$

$$\Delta y = \beta\Delta(\ln x)$$

$$lny = \beta x + u$$

x 变化 1 单位,y 变化 100β%

$$\Delta(\ln y) = \beta\Delta x = \beta \cdot 1 \Longrightarrow \frac{\Delta y}{y} \approx \beta$$

一般认为，第三种变化只改变截距，不改变系数。

证明

$$\begin{aligned}\log(y) &= \alpha + \beta x + u\\ \log(c) + \log(y) &= \log(c) + \alpha + \beta x + u\\ \log(cy) &= [\log(c) + \alpha] + \beta x + u\end{aligned}$$

但是实际情况由于包含 0 值，非更加复杂。

命题 3.1

推荐扩展阅读计量经济学：Log (y+1) 的转化是否可靠

3.6 匹配

计量自诩为因果分析的关键就是**可比性**。实验组和对照组的分析是最符合直觉的因果分析。

FWL 定理 (3.1.1) 告诉了我们为何 OLS 加入**控制变量**是保持其他变量不变。

当连续变量变为虚拟 01 变量时，这种分析就变为了因果效应的实验组 (d=1) 和对照组 (d=0)。这种变化将在9.1说明。

因此——加入控制变量-寻找匹配的反事实-寻找对照组。其实本质是一件事：**找到可比的对象说明因果关系**。核心问题都在于匹配、控制变量的选择是否具有内生性。

综上，同样的变量，是选择对样本进行匹配，还是直接加入控制变量，实际上和工具变量的使用一样，是无偏性和精准性的取舍问题。

3.7 多重共线性

- 多重共线性有个程度的问题。多重共线性既包括完全的线性关系, 又包括不完全的线性关系。
- 不完全多重共线性并不违背古典假定 MLR.3。
- 不完全多重共线性下,OLS 估计量仍是 BLUE 的。
- 多重共线性指解释变量间的线性关系。无多重共线性只排除解释变量间的线性关系, 不排除相互之间的非线性关系。(例子：p68 思考题 3.3)
- 不完全多重共线性不是计量经济学必须要解决的问题, 只有完全多重共线性才需要采取方法解决。
- 如果我们关注的变量和其它变量之间不存在多重共线性, 则可忽视多重共线性问题

。

笔记 完全多重共线性，指的是不能是线性共线关系，但是不排除相互的非线性关系。例如关系为 $\frac{x_1}{x_2}$

3.7.1 多重共线性与估计量

定义 3.2 (完全多重共线性)

对于解释变量 $1, x_1, x_2, ..., x_k$，如果存在不全为零的数 $\lambda_1, \lambda_2, \lambda_3, ..., \lambda_{k+1}$，使得：

$$\lambda_1 + \lambda_2 \boldsymbol{x_1} + \lambda_3 \boldsymbol{x_2} + \cdots + \lambda_{k+1} \boldsymbol{x_k} = 0$$

其中，$\boldsymbol{x_i} = (x_{1i}, x_{2i}, ..., x_{ni})'$ 为 n 维列向量。则称解释变量 $1, x_1, x_2, ..., x_k$ 之间存在完全多重共线性。

核心式子：

$$\hat{\beta}_j = \frac{\sum \widehat{\gamma}_{ij} y_i}{\sum \widehat{\gamma}_{ij}^2}, j = 1, 2, \cdots, k$$

详细介绍请回顾3.3章。

完全多重共线无法估计。

不完全多重共线:$\hat{\beta}_j = \frac{\sum \hat{\gamma}_{ij} y_i}{\sum \hat{\gamma}_{ij}^2}$

解释变量间线性无关：$\hat{\beta}_j = \frac{\Sigma(\hat{x}_{ij} - \bar{x}_j) y_i}{\Sigma(\hat{x}_{ij} - \bar{x}_j)^2}$

从不完全多重共线到解释变量间线性无关，本质差异是 x 对完全无关的解释变量回归相当于对常数项回归。

例题 3.1 共线性的判断

1、对于式子

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

(1)$x_1 + x_2 = 0$

(2)$x_1 + x_2 = a(a \neq 0)$

1、对于式子

$y = \beta_1 x_1 + \beta_2 x_2 + u$

(3)$x_1 + x_2 = 0$

(4)$x_1 + x_2 = a(a \neq 0)$

判断（1）-（4）每个式子是否完全共线。

解 (1)-(3) 都完全共线，(4) 则不会。

参考定义3.2。关键是回归方程的解释变量部分是否存在不完全为 0 的系数组合，使得式子（是否有常数项取决于回归方程的设定）为 0，只有（4）不存在，因为没有截距项。其他都是能找到对应的系数组合满足条件。

证明 完全多重共线性证明 假设存在共线关系：$x_1 = k \cdot x_2$。

会得到两个式子：

$$(k \cdot x_2) + x_2 = a$$

$$(k+1)x_2 = a$$

解的 $x_2 = \frac{a}{k+1}$ 只能是一个常数，由于线性关系，那么 x_1 也是常数，矛盾。

注 如果不想采用反证法，那么就用线性代数证明对应的矩阵 $(X'X)$ 是可逆的（即其行列式不为零）。

笔记 （2）和（4）的本质差异是（2）的共线性是截距、x_1 和 x_2 的共线。由于（4）没有截距，其实是判断 x_1 和 x_2 的共线。在没有截距项的情况下，完全共线性的定义变得更严格：一个解释变量必须是另一个解释变量的常数倍，但这点在 (4) 中并不成立。

以下是例题3.1的证明代码

Listing 3.2: 多重共线性-例题代码测试

```
clear
set obs 100 // 设置样本量 100
// 生成 x1
gen x1 = rnormal(0,1) // 正态随机数
// 情形 (1) 和 (3): x1 + x2 = 0
gen x2_case1 = -x1
gen x2_case3 = -x1
// 情形 (2) 和 (4): x1 + x2 = a (假设 a=5)
local a = 5
gen x2_case2 = `a' - x1
gen x2_case4 = `a' - x1
// 生成误差项 u
gen u = rnormal(0,1)
// 生成 y
// 情形 (1) 和 (2): 有截距
gen y_case1 = 2 + 3*x1 + 4*x2_case1 + u
gen y_case2 = 2 + 3*x1 + 4*x2_case2 + u
// 情形 (3) 和 (4): 无截距
gen y_case3 = 3*x1 + 4*x2_case3 + u
gen y_case4 = 3*x1 + 4*x2_case4 + u
// 回归测试
// 情形 (1)
display "===== Case 1: y = b0 + b1*x1 + b2*x2, x1+x2=0 ====="
reg y_case1 x1 x2_case1
// 情形 (2)
display "===== Case 2: y = b0 + b1*x1 + b2*x2, x1+x2=a ====="
reg y_case2 x1 x2_case2
// 情形 (3)
display "===== Case 3: y = b1*x1 + b2*x2, x1+x2=0 ====="
reg y_case3 x1 x2_case3, noconstant
// 情形 (4)
display "===== Case 4: y = b1*x1 + b2*x2, x1+x2=a ====="
reg y_case4 x1 x2_case4, noconstant
```

3.8 异方差

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

在计量前四个假定下（不包含同方差假定）。

此时误差包含异方差，写为 $\mathrm{Var}(u_i \mid x_i) = \sigma_i^2$

此时可以得到以下方差估计

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma_i^2}{\mathrm{SST}_x^2}$$

其中

$$\mathrm{SST}_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

。

在多元回归中

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

$$\widehat{\mathrm{Var}(\hat{\beta}_j)} = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{\mathrm{SSR}_j^2}$$

其对应的平方根就是异方差-稳健标准误。其中 $\mathrm{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$。

在讨论共线性时，可以使用 $\mathrm{SST}_j(1 - R_j^2)$ 替代 SSR_j。

注[不同层次] 式子的 SST 和 SSR 针对的样本 x 和样本 y，而其中的 R_j^2 是解释变量对其他控制变量回归的拟合优度。

此时 $\mathrm{SST}_j(1-R_j^2)$ 可以理解为 x_j 剔除了其他变量后的有效平方和，对应其独立贡献。因此才是 $1-r^2$，这部分是 x_j 没被其他控制变量解释的，独立的贡献。

3.8.1 GLS

命题 3.2 (异方差的后果)

异方差的情况在：

- 高斯-马尔可夫定理不再成立。OLS 估计量不再是 BLUE
- t 检验和 F 检验不再有效
- OLS 估计量仍然无偏

广义最小二乘目的是在异方差和自相关的情况在依旧是最优无偏估计。

同方差模型下：

$$y = X\beta + u, \quad E[u] = 0, \quad \mathrm{Var}(u) = \Omega$$

gls 则是给其加权正定矩阵

$$\tilde{y} = \Omega^{-1/2}y, \quad \tilde{X} = \Omega^{-1/2}X, \quad \tilde{u} = \Omega^{-1/2}u$$

$$\tilde{y} = \tilde{X}\beta + \tilde{u}, \quad \mathrm{Var}(\tilde{u}) = I_n$$

gls（广义最小二乘）其实是 wls（加权最小二乘）的一种推广。或者说：wls 是 gls 的一种特殊形式。

权重取决于假设：

如果**只有异方差**，误差独立但方差不同。

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

其中对角线方差并不完全相等。

如果**只存在自相关**，对角线方差相同，但非对角线协方差不为 0。

$$\Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

其中 $\sigma_1 = \sigma_2 = \cdots = \sigma_n$

例题 3.2gls 和 wls 人口抽样调查时有权重，将权重参与回归就是加权最小二乘，处理的只是异方差问题，因为其权重矩阵为对角矩阵吗，为特殊的 gls。

标准 re 模型为同方差模型，gls 只处理自相关。但是如果数据实际上存在异方差，构建的矩阵也能处理异方差和自相关。

fe 也可以看作一种特殊的 gls，通过特殊的权重（组间变换矩阵）处理方程。异方差就是使用稳健标准误。

3.8.2 stata 实现

- 自带命令:xtreg y x,fe
- 外部命令:xtivreg2 y x, fe/fd(安装 xtivreg2 ssc install xtivreg2)
- 高维固定效应:reghdfe y x,absorb(year id)

注[stata 检验] xtreg 中的选项 robust 是异方差-自相关稳健标准误（异方差和自相关的差别可回顾3.8）。

也就是说，下面三种命令是等价的：

xtreg y x, fe robust=xtreg y x, fe cluster(id)=xtivreg2 y x,fe vce(cluster id)

下面三种则都是异方差稳健标准误：

xtivreg2 y x, fe vce(robust)=reghdfe y x, absorb(year id) vce(robust)=reg y x i.year i.id, robust

第四章 渐进分布假设检验

4.1 分布基础知识

正态分布：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty$$

记作 $X\sim N(\mu,\sigma^2)$
数字特征：$E(x)=\mu$, $\mathrm{Var}(x)=\sigma^2$。
正态分布标准化：
若 $X\sim N(\mu,\sigma^2),\eta=\frac{x-\mu}{\sigma}$, 则 $\eta\sim N(0,1)$。

定理 4.1 (正态分布的性质)

性质 1：线性组合。$X\sim N(\mu,\sigma^2)$ 则 $aX+b\sim N(a\mu+b,a^2\sigma^2)_1$。
性质 2：独立同分布正态随机变量的任意线性组合都是正态分布。
推论；$Y_j\sim N(\mu,\sigma^2)$ 则 $\overline{Y}\sim N(\mu,\sigma^2/n)$

χ^2 分布：
令 $Z_i,i=1,2,...,n$ 为独立随机变量, 且都服从标准正态分布。定义一个新随机变量为 Z_i 的平方和：

$$X=\sum_{i=1}^{n}Z_i^2$$

那么,X 即所谓具有 n 个自由度的 χ^2(chi-square) 分布, 记
为 $X\sim\chi_n^2$

定理 4.2 (χ^2 分布的性质)

χ^2 分布具有可加性, 即当 Y 和 Z 相互独立,
$Y\sim\chi_n^2,Z\sim\chi_m^2$, 则 $Y+Z\sim\chi_{n+m}^2$

笔记 卡方分布只满足相加性，不存在减。因为卡方分布区间为 0 到正无穷，相减可能为负。

***t* 分布**
若当 X 和 Y 相互独立,$X\sim N(0,1),Y\sim\chi_n^2$, 则随机变量

$$T=\frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布, 记为 $T\sim t_n$。

t 分布的概率密度函数有一个类似于标准正态分布的形状, 只是它更散开一些, 因而尾端有较大的面积。

性质：当 $n\to\infty$ 时,t 分布近似于标准正态分布。

笔记 给定同样的区间概率 α,t 统计量比正态分布的分位数更大（绝对值）。因为 t 检验更瘦一点，也意味着图像长尾。

F 分布
令 $X_1\sim\chi_{k_1}^2$, $X_2\sim\chi_{k_2}^2$, 且 X_1 和 X_2 相互独立, 则随机变量

$$F=\frac{X_1/k_1}{X_2/k_2}$$

服从自由度为 (k_1,k_2) 的 F 分布, 记为 $F\sim F_{k_1,k_2}$。

笔记 在给定相同置信水平 $1-\alpha$ 时，t 分布的临界值（绝对值）通常比标准正态分布的临界值更大：

$$|t_{\nu,\alpha/2}|>|z_{\alpha/2}|,\quad \nu<\infty$$

分位数

随机变量的分布函数为 $F(X)$, 实数 α 满足￥$0<\alpha<1$ 时：

α 分位数是使 $P\{X<x_\alpha\}=F(x_\alpha)=\alpha$ 的数 x_α

上分位数：$P\{X>\lambda\}\}=1-F(\lambda)=\alpha$ 的数 λ

双侧分位数：

$P\{X<\lambda_1\}\}=F(\lambda_1)=\alpha_1$

$P\{X<\lambda_2\}\}=1-F(\lambda_2)=\alpha_2$

若 x 概率密度函数关于 y 轴对称，则 $\lambda_2=-\lambda_1$。

4.2 推断统计假设

假定：误差正态分布

以 X 为条件,u_t 服从独立同分布 $Normal(0,\sigma^2)$。换而言之, 给定 X,u 服从均值为 0 和方差一协方差矩阵为 $\sigma^2\mathbf{I}_n$ 的多元正态分布 $\mathbf{u}\sim\text{Normal}(0,\sigma^2\mathbf{I}_n)$。

例如对于每一个 t,u_t 独立与解释变量, 在时间序列中就是严格外生性假定。

定理 4.3

在经典线性模型假定下, 以 X 为条件,$\hat\beta$ 服从均值为 β 和方差一协方差矩阵为 $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ 的多元正态分布。

通过公式可以看作分布的传递

$$\widehat{\beta}_j=\frac{\sum\widehat{\gamma}_{ij}y_i}{\sum\widehat{\gamma}_{ij}^2}=\beta_j+\frac{\sum\widehat{\gamma}_{ij}u_i}{\sum\widehat{\gamma}_{ij}^2},(j=1,2,\cdots,k)$$

$$u_i\to y_i\to\hat\beta_j$$

定理 4.4

在经典线性模型假定下, 以 X 为条件, 有 t 分布统计量 $(\hat\beta_j-\beta_j)/\text{se}(\hat\beta_j)\sim t_{n-k-1},j=0,1,\cdots,k$

证明 这里主要通过正态分布假设提取 **t** 分布的变形**,** 体会转换思路**,** 寻找其他分布检验量是同样的思想。

t 分布的转化思路是——正态分布转化为 χ 分布, 证明要素相互独立, 制造 t 分布。

由于 $\hat\beta$ 服从正态分布, 得到标准化 $(\hat\beta_j-\beta_j)/\text{sd}(\hat\beta_j)\sim Normal(0,1)$

定理3.3.2证明了 $\text{Var}(\hat\beta\mid\mathbf{X})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$,

对应的标准差就是 $\text{sd}(\hat\beta_j)=\sigma\sqrt{c_{jj}}$。$c_{jj}$ 代表 $(\mathbf{X}'\mathbf{X})^{-1}$ 对角线上的第 j 个元素。

接下来继续使用误差正态分布的假定 $\mathbf{u}\sim\text{Normal}(0,\sigma^2\mathbf{I}_n)$, 得到 $\mathbf{u}/\sigma\sim\text{Normal}(0,\mathbf{I}_n)$

基于对称幂矩阵 M(n*n 的矩阵) 的性质, 将正态分布转化为 χ 分布。

$(n-k-1)\hat\sigma^2/\sigma^2=(\mathbf{u}/\sigma)'\mathbf{M}(\mathbf{u}/\sigma)\sim\chi^2_{n-k-1}$

接下来为组合 t 检验量, 需要保证 $\hat\beta$ 和 $\hat\sigma$ 相互独立。

注 多元正态分布关于独立性的性质。

如果 $y \sim Normal(0, \sigma^2 \mathbf{I}_n)$, A 是一个 $k \times n$ 非随机矩阵, 而 B 是一个 $n \times n$ 对称幂等矩阵, 那么 Ay 和 y′By 独立的充分必要条件是 AB=0。

可以看到估计量 $\hat{\beta} = \beta + (X'X)^{-1}X'u$ 和 $\hat{\sigma}^2 = \mathbf{u}'\mathbf{M}\mathbf{u}/(n-k-1)$ 相乘, 因为消灭矩阵的性质 $MX = 0$, 因此, 相乘也为 0, 则相互独立。

$\hat{\sigma}^2$ 是 Mu 的函数, 自然 $\hat{\beta}$ 和 $\hat{\sigma}$ 也相互独立。

于是组合出 t 分布 $(\hat{\beta}_j - \beta_j)/\mathrm{se}(\hat{\beta}_j) = [(\hat{\beta}_j - \beta_j)/\mathrm{sd}(\hat{\beta}_j)]/(\hat{\sigma}^2/\sigma^2)^{1/2} \sim t_{n-k-1}$

4.3 渐进分布

之前研究了有效性 (涉及期望), 有效性 (涉及方差), 接下来就是一致性。当样本容量不断扩大时, 估计趋于一致, 也就是渐进。**也就是当我们拥有大样本时，大样本自身的优秀性质可以使我们放松一些原来的假设。**

定理 4.5

满足线性弱相关、无完全共线性、零条件均值, 则 OLS 估计量是一致的 $plim\hat{\beta}_j = \beta_j$。

证明

OLS 估计量如下：

$$\widehat{\beta} = \left(\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{y}_t\right) = \left(\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(\sum_{t=1}^{n} \mathbf{x}_t'\left(\mathbf{x}_t\beta + u_t\right)\right)$$

$$= \boldsymbol{\beta} + (\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t)^{-1}(\sum_{t=1}^{n} \mathbf{x}', u_t)$$

$$= \boldsymbol{\beta} + (n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t)^{-1}\left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'u_t\right)$$

中间提取 n 是为了使用大数定理。

定理 4.6 (大数定理)

弱大数定律 (weak law of large numbers,WLLN): 若 $G\times 1$ 维独立同分布的随机向量序列 $\{\boldsymbol{w}_i : i = 1, 2, \cdots, N\}$ 满足, $\mu_w = E(w_i), E(|w_{ig}|) < \infty (g = 1, \cdots, G)$, 那么,

$$N^{-1}\sum_{i=1}^{N} w_i \xrightarrow{p} \mu_w$$

基于大数定理, 得到

$$n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t \xrightarrow{p} \mathbf{A} \text{ 和 } n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'u_t \xrightarrow{p} \mathbf{0}$$

第一个式子使用了无完全共线性, 说明矩阵 A 是 $(k+1)\times(k+1)$ 非奇异矩阵；后者利用了零条件均值。

对于非奇异矩阵 A, 有

$$\left(n^{-1}\sum_{t=1}^{n} \mathbf{x}'{}_t\mathbf{x}_t\right)^{-1} \xrightarrow{p} \mathbf{A}^{-1}$$

综上可以发现：

$$plim(\hat{\beta}) = \beta + A^{-1} \cdot 0 = \beta$$

定理 4.7

满足线性弱相关、无完全共线性、零条件均值、同方差、无序列自相关, 则 OLS 估计量是渐进正态分布。且通常都 OLS 标准误、t 统计量、F 统计量、LM 统计量是渐进有效的。

计量经济学导论这部分使用的时间序列, 条件分别如下 TS：

1. 线性和弱相关 (例如引进滞后变量)
2. 无完全共线性
3. 零条件均值
4. 同方差
5. 无序列相关

条件 1-3 即可得到条件 4——OLS 估计量一致性, 且 OLS 标准误、t 统计量、F 统计量、LM 统计量是渐近有效的。条件 1-5 即可证明 OLS 是渐进正态分布的。且 OLS 标准误、t 统计量、FM 统计量和 LM 统计量是渐进有效的。

证明 继续使用大数定理的构造：

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}-\beta\right) &= \left(n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t\right)^{-1}\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t\right) \\ &= \mathbf{A}^{-1}\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t\right) + o_p(1)\end{aligned}$$

$o_p(1)$ 是依概率收敛与 0 的余项, 等于 $\left[\left(n^{-1}\sum_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t\right)^{-1} - \mathbf{A}^{-1}\right]\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{x}_t'u_t\right)$。

接下来使用中心极限定理

定理 4.8 (中心极限定理)

中心极限定理 (central limit theorem): 若 $G \times 1$ 维独立同分布的随机向量序列 $\{\boldsymbol{w}_i : i = 1, 2, \cdots, N\}$ 满足,$E(\boldsymbol{w}_i) = \mathbf{0}$, $E(w_{ig}^2) < \infty (g = 1, \cdots, G)$, 那么,

$$N^{-1/2}\sum_{i=1}^{N}\boldsymbol{w}_i \xrightarrow{d} Normal(0, \boldsymbol{B})$$

其中,$B = Var(w_i) = E(w_i w_i')$, B 是 $G \times G$ 维方差协方差矩阵。

$(n^{-1/2}\sum_i^n \mathbf{x}', u_i)$ 收敛于多元正态分布, 则依概率有界。两项之积收敛于 0。

$\sqrt{n}\left(\hat{\beta}-\beta\right)$ 从 $\mathbf{A}^{-1}\left(n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i'u_i\right)$ 得到了渐近分布。

太累了**...** 接下来使用 **ai** 帮助复制原文。

根据中心极限定理,$(n^{-1/2}\sum_{i=1}^{\pi}\mathbf{x}_i'u_i)$ 服从均值为零和 $(k+1)\times(k+1)$ 方差一协方差矩阵为 B 的一个渐近正态分布。于是,$\sqrt{n}(\hat{p}-\hat{p})$ 便服从均值为零和方差一协方差矩阵为 $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ 的渐近多元正态分布。

我们现在证明了, 在假定 TS4′ 和假定 TS5′ 下,B=σ^2**A**。(这里一般表达式也有用处, 因为它成为第 12 章讨论的 OIS 之异方差一稳健和序列相关稳健标准误的基础。) 首先, 在假定 TS 5′ 下, 对 $t \neq s, \mathbf{x}'u$, 和 $\mathbf{x}'u$, 无关, 为什么? 为简单起见, 假设 $s < t$。于是, 根据迭代期望法则,$\mathbb{E}(\mathbf{x}_i'u_iu_i\dot{\mathbf{x}}_i) = \mathbb{E}[\mathbb{E}(u_iu_i|\mathbf{x}_i'\mathbf{x}_i)\mathbf{x}_i'\mathbf{x}_i] = \mathbb{E}[\mathbb{E}(u_iu_i|\mathbf{x}_i'\mathbf{x}_i)\mathbf{x}_i'\mathbf{x}_i] = \mathbb{E}[0 \cdot \mathbf{x}, '\mathbf{x},] = 0$。协方差为零便意味着和的方差等于方差之和。但 $\mathrm{Var}[\mathbf{x}_i'u_i] = \mathbb{E}(\mathbf{x}_i'u_iu_i\mathbf{x}_i) = \mathbb{E}(u_i^2\mathbf{x}_i'\mathbf{x}_i)$。根据迭代期望法则,$\mathbb{E}(u_i^2\mathbf{x}_i'\mathbf{x}_i) = \mathbb{E}[\mathbb{E}(u_i^2\mathbf{x}_i'\mathbf{x}_i|\mathbf{x}_i)] = \mathbb{E}[\mathbb{E}(u_i^2|\mathbf{x}_i)\mathbf{x}_i'\mathbf{x}_i] = \mathbb{E}(\sigma^2\mathbf{x}_i'\mathbf{x}_i) = \sigma^2\mathbb{E}(\mathbf{x}_i'\mathbf{x}_i) = \sigma^2\mathbb{A}$, 其中我们利用了在假定 TS. 3’ 和假定 TS.4’ 下 $\mathrm{E}(u_i^2|\mathbf{x}_i) = \sigma^2$ 的结论。这就证明了 B= σ^2**A**,

所以在假定 TS.1′ 至假定 TS.5’ 下, 我们有

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

根据方程 $\sqrt{n}(\hat{\beta}-\beta) \overset{a}{\sim} \mathrm{Normal}(0, \sigma^2 \mathbf{A}^{-1})$, 我们把 $\hat{\beta}$ 视为近似均值为 β 和方差为 $\sigma^2 A^{-1}/n$ 的正态分布。这里除以样本容量 n 在预料之中：$\hat{p}$ 的近似方差一协方差矩阵以速度 $1/n$ 收敛于零。当我们将 σ^2 代之以其一致估计量 $\hat{\sigma}^2$ =SSR/$(n-k-1)$, 并将 A 代之以其一致估计量 $n^{-1}\sum_{i=1}^{n}\mathbf{x}_i'\mathbf{x}_t = \mathbf{X}'\mathbf{X}/n$, 我们便得到 $\hat{\beta}$ 的渐近方差估计量。

$$\widehat{\mathbf{A}\mathrm{Var}(\hat{\boldsymbol{\beta}})} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

总之, 我们证明了, 在假定 TS.1′ 至假定 TS.5′ 下 (假定 TS.1 至假定 TS.5 是其特殊情形), 通常的标准误和 ι 统计量都是渐近有效的。欲求单个假设检验的临界值和 p 值, 使用通常的 t 分布完全合法。有意思的是, 在第 11 章的一般背景下, 假定误差的正态性「比如给定 $\mathbf{x}_i, u_{i-1}, \mathbf{x}_{i-1}, \cdots, u_i, \mathbf{x}_i$ 下,u_i 服从 Normal(0, σ^2)] 不一定有帮助, 因为在这种正态性假定下,ι 统计量一般都不具有精确的 t 分布。我们若不假定解释变量的严格外生性, 欲求精确的分布结论, 即便不是不可能, 也会极其困难。

如果我们对上述论证略加修改, 便可推导出一个异方差一稳健的方差一协方差矩阵。关键在于, 我们必须单独估计 $\mathbb{E}(u_i^2\mathbf{x}_i'\mathbf{x}_i)$, 因为它不再等于 $\sigma^2\mathbb{E}(\mathbf{x}_i'\mathbf{x}_i)$。但如果 $\hat{u}_i$ 是 OLS 残差, 则一个一致估计量便是

$$(n-k-1)^{-1}\sum_{i}^{n}\hat{u}_t^2\hat{\mathbf{x}}_t'\mathbf{x}_t$$

其中, 除数 $n-k-1$ 而非 n 是调整后的自由度, 通常有助于估计量的有限样本性质。我们利用方程中的表达式, 便得到

$$\widehat{\mathrm{AVar}(\hat{\boldsymbol{\beta}})} = [n/(n-k-1)]\,\mathbf{X}'\mathbf{X}^{-1}\left(\sum_{i=1}^{n}\hat{u}_i^2\mathbf{x}_i'\mathbf{x}_i\right)(\mathbf{X}'\mathbf{X})^{-1}$$

这个矩阵主对角线元素的平方根, 正是我们在 8.2 节在纯粹横截面情形中得到的异方差一稳健的标准误。我们还能将 12. 5 节得到的序列相关一 (和异方差一) 稳健标准误进行矩阵形式的扩展, 但由于序列相关, 取代 (E.25) 的矩阵相当复杂。比如参见 Hamilton(1994,Section 10-5)。

练习 4.1 渐进与一致 渐进和一致的关系？

证明 [一致性和渐进性关系] 关系

证明一致性的工具是大数定理，证明渐进性的工具是中心极限。

渐进（asymptotic）不一定一致（consistent）。但是，一致性通常是渐近性质的一种特殊形式。因为渐进是一个宽泛的概念，是分布趋于一致，例如渐进正态、渐进有效、渐进无偏。

例如：

$$\hat{\theta}_n = \begin{cases} \theta_0 + 1 \text{概率} 1/2 \\ \theta_0 - 1 \text{概率} 1/2 \end{cases}$$

其渐进正态但不一致。只有当渐进分布的中心是真值时，渐进正态才能趋于一致。

4.4 OLS 条件估计总结

在多元 OLS 估计下, 有以下条件 MLR1-6

1. 线性于参数

2. 随机抽样
3. 不存在完全共线性
4. 零条件均值
5. 同方差
6. 正态性：

条件 1-4 下,OLS 估计量是无偏的 (unbiasedness)。截距和斜率估计量也是一致的估计量。

条件 1-5 下 (高斯马尔科夫假设),OLS 估计量是最优线性无偏估计量 (best linear unbiased estimator-BLUE)。也是渐进有效的。$\hat{\beta}$ 符合渐近正态分布,$\hat{\sigma}$ 是 σ 的一致估计。

条件 1-6 下,OLS 估计量是最小方差无偏估计量 (MVUE)(相比于 BLUE 不再要求线性)。

样本容量为任意 n 有限时, 这些性质都成立。于是接下来研究样本容量任意大时的情形, 这时候我们倾向于寻找具有一致性的情况。

也就是在大样本条件下, 可以放松条件 6, 前提是 5 的同方差必须满足。

注 条件 6：总体误差独立于解释变量, 服从均值为 0 和方差为 σ 的正态分布。这个条件只是为了能使用 t 统计量和 F 统计量做精确的推断。但大样本能满足渐近正态分布, 因此可以放松条件 6。

例题 4.1 有限样本有偏, 大样本一致 假设 Z 的真值为 0, 一个随机变量 X 以 (n-1)/n 的概率取值为 Z, 而以 $1/n$ 的概率取值为 n。那么,X 的期望为 1, 也就是：

$$E(X) = Z\frac{n-1}{n} + n\frac{1}{n} = 1$$

记 $plim(x)$ 为 n 趋向无穷大时 x 的取值, 则有：$plim(x) \neq z = 0$

例题 4.2 估计量无偏但不一致 依然假设 Z 的真值为 0, 一个随机变量 X 以 0.5 的概率取 0.5, 而以 0.5 的概率取-0.5, 那么 X 的期望为 0, 也就是说,X 是 Z 的无偏估计量。但是,X 总是在 $X = 0$ 这条线上下摆动, 当 n 趋向无穷大时, 它的方差并不会趋于 0。因此,X 并不是 Z 的一致估计量, 也就是说 X 不具备一致性。

无偏估计量未必是一致的, 但是那些当样本容量增大时方差会收缩到零的无偏估计量是一致的。

4.5 假设检验

笔记 假设检验的思想

在某种条件（原假设）下，在一次抽样中，大概率事件出现被认为是合理的，而小概率事件被认为基本不会发生，如果小概率事件竟然发生了，认为（原假设）是不合理的。

在事先作出的某种原假设成立的条件下，利用样本构造适当统计量（一次抽样的结果），并确定统计量的抽样分布。给定显著性水平，构造一个小概率事件。如果在一次抽样中该小概率事件竟然发生，就认为原假设不真实，从而拒绝原假设，不拒绝备择假设。反之，如果大概率事件发生，则不拒绝原假设。

$$\text{分布} \rightarrow \text{检验统计量} \rightarrow \begin{cases} \text{假设检验：估计值与真实值的差异衣服显著} \\ \text{置信区间：真实值在什么区间内} \end{cases}$$

原理：误差大到一种“可能性很小”的程度, 也就是容忍度——5%、1%。

于是我们构造小概率事件——第一类错误和第二类错误。

原假设 $H_0 : \theta = 0.42$

备择假设 $H_0 : \theta > 0.42$

单侧双侧、分位数、几种常用分布。

经济学检验一般检验异于 0, 因为经济含义中, 系数为 0 则无影响。

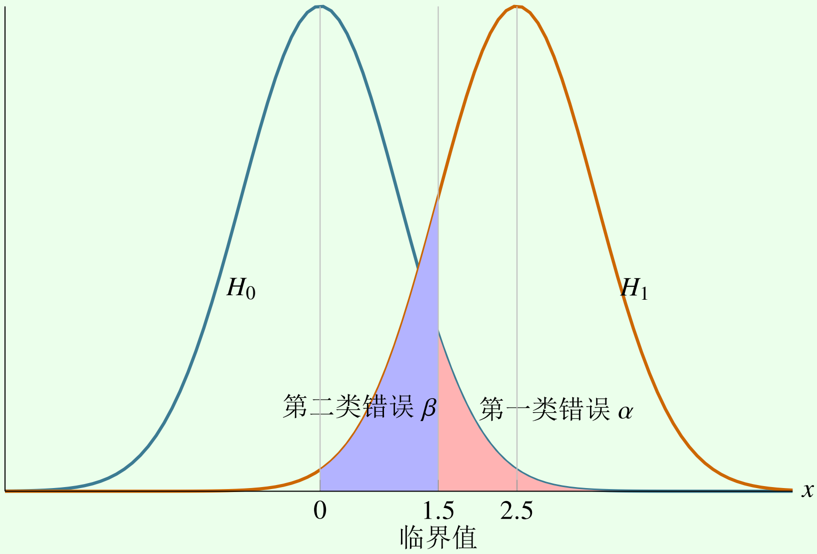
第一类错误： 原假设为真时拒绝 $\alpha = \text{P}(\text{拒绝 H}_0 \mid \text{H}_0)$, 也就是显著性检验。

第二类错误： 原假设为错时接受 $\beta = \text{P}(\text{接受 H}_0 \mid \text{H}_1)$

注 **原假设设置原则**

将想要拒绝的假设放在原假设。(经济学就是系数为 0)

将证伪成本最小的假设放在原假设。(例如疑罪从无, 原假设是有罪；外星人存在, 是存在性问题, 只需要找到例子即可, 证明成本最小)



α 和 β 关系为此消彼长, 每次只能保证一个都精准度, 想要同时增加只能增大样本容量。临界值为给定数值, 因此 H_0 和 H_1 共享一个临界值。

临界值的确定就是容忍度的设置：决定显著性水平, 给定 α, 临界值 c 由 H_0 被假定正确时的分布决定, 因此 H_1 和 H_0 共享一个临界值。

检验功效：$\pi = 1 - \mathrm{P}(\text{拒绝 } \mathrm{H}_0 \mid \mathrm{H}_0)$

有检验统计量 t

$\mathrm{H}_0 : \mu = \mu_0$的$t$统计量($t$statistic):t

$t = \sqrt{n}(\bar{y} - \mu_0)/s = (\bar{y} - \mu_0)/\mathrm{se}(\bar{y})$

t_{n-k+1} 统计量用 $\bar{y}$ 的标准误 se 度量了 $\bar{y}$ 到 μ_0 的距离。

$T = \sqrt{n}(\bar{Y} - \mu_0)/S$

$\mathrm{se}(\bar{y}) = s/\sqrt{n}$ 是 $\bar{y}$ 的标准误

P 值双侧 $\mathrm{P}(\mid T_{n-1} \mid > \mid t \mid) = 2\mathrm{P}(T_{n-1} > \mid t \mid)$

$p = P(T > 1.52|H_0) = 1 - \Phi(1.52) = 0.065$

T 值是渐进近似检验, 大样本收敛于正态分布。对于非正态分布样本, 检验量实际是渐近的统计检验量。因此大样本可以使用 z 检验。

计量经济学导论第六版 632 页 E.3。

注 因为是样本拟合, 实际上是使用标准误 (样本方差) 替换总体方差公式。因此小样本是 t 检验, 大样本基于渐进分布可以使用 z 检验。大样本依旧可以使用 t 检验, 因为 t 分布相对 z 分布更加瘦, 是更加保守的推断, 如果拒绝了原假设, 那么 z 检验依旧成立。

$T = \sqrt{n}(\bar{Y} - \mu_0)/S \overset{a}{\sim} \mathrm{Normal}(0, 1)$

P 值：拒绝原假设所给定的最小显著性水平；不能拒绝原假设的最大显著性水平

双测检验能推导出单侧检验

双侧 $2P < \alpha$

单侧 $P < \alpha$

$P < 2P < \alpha$

双侧检验：

$$P\left(|T| > \left|t_{\widehat{\beta}_j}\right| = \left|\frac{\widehat{\beta}_j}{se(\widehat{\beta}_i)}\right|\right) = p$$

右侧检验：

$$P\left(T > t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j}{se(\widehat{\beta}_j)}\right) = p$$

左侧检验：

$$P\left(T < -t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j}{se(\widehat{\beta}_j)}\right) = p$$

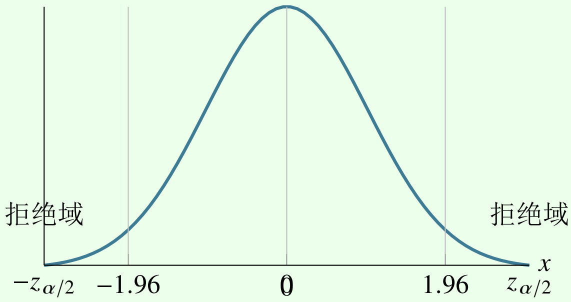
定义 4.1 (p 值)

正确表述：在 $x\%$ 的显著性水平上统计显著。完整表述为“在 $x\%$ 的显著性水平上拒绝 H_0 接受 H_1”。♣

笔记 理解的注意点

- p 值不是“原假设为真的概率”
- p 值不是“实验结果由偶然造成的概率”。
- p 值不是“显著性水平 α

如下图所示, 拒绝域的面积为 P 值。



经济显著性：估计值的大小。

例题 4.3 统计显著

$$y = \beta_0 + \beta_1 x$$

将 x 的单位从元换成万元,β 统计显著性不变, 数值增大一万。

统计性显著：给定显著性水平上达到显著。

例题 4.4 经济显著

$$y = \beta_0 + \beta_1 x$$

方法一:$\frac{\hat{\beta}}{\bar{y}} vs$ 其他研究

方法二：β 系数

4.6 t 检验

对于 ols 来说，对应的假设是 $\beta = 0$, 构造 t=$\frac{\beta-0}{\sqrt{var\beta}}$

4.7 置信区间

$$P(\widehat{\beta}_j - \delta \le \beta_j \le \widehat{\beta}_j + \delta) = 1 - \alpha$$

此时 $1-\alpha$ 为置信水平，$\widehat{\beta}_j + -\delta$ 为置信上下限。

定义 4.2 (置信区间)

若反复抽样多次 (每次的样本容量相同)，每个样本值确定一个区间 $(\widehat{\beta}_j - \delta, \widehat{\beta}_j + \delta)$。每个这样的区间要么包含 β, 要么不包含 β。按照伯努利大数定理，在这么多区间中，包含的区间约占 $100(1-\alpha)\%$, 不包含的区间约占 $100\alpha\%$。

给定参数估计量的统计分布，总能找到包含真实 $\hat{\beta}_j$ 的概率为 $1-\alpha$ 的区间（存在无数个）。但是我们希望找到最短的区间，也就是置信区估计精度最高。

对于正态或者 t 分布的统计量 $\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)}$，当 n 固定时，双侧对称点区间最短：

$$P\left(-t_{\alpha/2} \le \frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \le t_{\alpha/2}\right) = 1 - \alpha$$

这意味着可以通过这个思想进一步检验 β 的线性组合：

$H_0{:}\beta_i = \beta_j$

可以直接检验 $t = \frac{\widehat{\beta}_i - \widehat{\beta}_j}{se(\widehat{\beta}_i - \widehat{\beta}_j)}$

也可以间接检验 $\theta_k = \beta_i - \beta_j$

4.8 F 检验

多个检验时可以使用 F 检验

有以下写法：

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$(排除性约束)

$H_1 : \beta_{1-3}$ 至少有一个不为 0。

笔记 为什么不能用 t 检验

为什么不能用 t 检验逐个检验 β_i。有可能单个不显著，联合起来显著；也可能单个显著，联合起来不显著。例如一正一负影响互相抵消。

$$H_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = 0$$

系数矩阵 $R\beta = r$

此时构造如下 F 检验量

$$F = \frac{SSR_r - SSR_{ur}}{\frac{SSR_{ur}}{n-k-1}} \sim F_{qn-k-1}$$

转化逻辑如下：对于多个约束，进而检验模型残差平方和（约束越正确，解释性越大）

$H_0 : SSR_r - SSR_{ur} = 0$ $H_1 : SSR_r > SSR_{ur}$(受到约束的模型残差平方和)

F 检验是一个右侧检验。因 R^2 随解释变量数量增加而增加，因此相减一定大于 0

例题 4.5 约束检验的扩展

同理，如果对于 $H_0 : \beta_1 = 1\beta_2 = 2\beta_3 = 3$ 的约束检验，

则是在回归中加入 $-x_1, -2x_2, -3x_3$。

证明 方程形式

如果 H_0 成立，有以下情况, 此时两个式子都是方差的无偏估计量:

$$\begin{cases} E(\frac{SSR_{ur}}{n-k-1}) = \sigma^2 \\ E(\frac{SSR_r}{n-k-1+q}) = \sigma^2 \end{cases}$$

H_0 成立影响的是两个式子的 σ^2 相等而非式子形式本身。

为什么能转换为 F 统计量? 分子分母分别是 χ 统计量。

$\frac{(n-k-1)\frac{SSR_{ur}}{n-k-1}}{\sigma^2} \sim \chi^2_{n-k-1}$

$\frac{(n-k-1)\frac{SSR_r}{n-k-1+q}}{\sigma^2} \sim \chi^2_{n-k-1+q}$

分子是两个卡方分布相减，依旧是卡方分布，原因是特殊的。

原因是 H_0 成立时，解释变量和残差（正交）不相关。

举例如下:

对于式子

$$y = \hat{\beta}_0 + \hat{\beta}_4 x_4 + \hat{u}^r$$

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{u}^{ur}$$

如果 H_0 成立，则 $\beta_1 = \beta_2 = \beta_3 = 0$

也就是说 $\hat{u}^r = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{u}^{ur}$

因此可以转换残差平方和 $\sum \hat{u}_i^{r2} = \sum(\hat{x}_i + \hat{u}_i{}^{ur})^2$

展开这个式子才会用到 $\hat{u}_i{}^{ur}\hat{x}_i = 0$

证明 矩阵形式

$$SSR_{ur} = \hat{u}'_{ur}\hat{u}_{ur} = y'(I-P)y$$

$$SSR_r = y'(I-P_r)y.$$

$$SSR_r - SSR_{ur} = y'(I-P_r)y - y'(I-P)y = y'(P-P_r)y.$$

此时 $(P-P_r)$ 是对称、幂等且秩为 q（正好对应被约束掉的自由度）, 且和 $(I-P)$ 正交。

问题 4.1 例子，下图是 stata 的检验结果，其中右上角 F 统计量的位置缺失了，尝试填写出来。

解 答案就是在左上角的小表格中，df 一列代表自由度，ss 一列除以自由度一列 df 就是 MS 一列。MS 一列上下相除就是结果都 F 统计量。也就是系数的联合检验。

定理 4.9 (F 检验的等效性)

$$F = \frac{(R^2 - R_*^2)/J}{(1-R^2)/(n-k)} = \frac{(SSE^* - SSE)/J}{SSE(N-K)} = \frac{(e'_* e_* - e'e)/J}{e'e/(n-k)}$$

4.9 渐进检验

关于渐进分布4.5。

```
      Source |       SS           df       MS      Number of obs   =     10146
-------------+----------------------------------   F(          ,         ) =
       Model |  100.813197         5  20.1626394   Prob > F        =    0.0000
    Residual |  1172.34673    10140  .115616048    R-squared       =    0.0792
-------------+----------------------------------   Adj R-squared   =    0.0787
       Total |  1273.15993    10145  .125496296    Root MSE        =    .34002

------------------------------------------------------------------------------
  lnweek_hour |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
lnfama_transfer |  -.0050192   .0010269    -4.89   0.000    -.0070321   -.0030062
       lnwage |   .0134061   .0033978     3.95   0.000     .0067459    .0200664
         male |   .0565741   .0070031     8.08   0.000     .0428466    .0703015
          age |  -.0035352   .0003905    -9.05   0.000    -.0043007   -.0027697
          edu |  -.0263465   .0009931   -26.53   0.000    -.0283931      .0244
        _cons |   4.138826   .0376791   109.84   0.000     4.064968    4.212685
------------------------------------------------------------------------------
```

图 4.1: F 检验

大样本用的怀特 wald 检验。

关于渐进性的讲义——计量经济学讲义

基于样本趋于无穷时的特点来研究。

笔记 [wald 检验] wald 检验被看作三大检验之一，是一种检验思想，用于无约束的检验。

要检验 $H_0 : R\beta = r$，其实就是检验线性组合：

$$R\hat{\beta} - r$$

由于大数定理，我们认为参数服从正态分布 $\hat{\beta} \sim N(\beta, \mathrm{Var}(\hat{\beta}))$，也就得到了对应检验量线性组合的分布

$$R\hat{\beta} - r \sim N(R\beta - r, R\,\mathrm{Var}(\hat{\beta})R')$$

WALD 检验量就是其标准化平方

$$W = (R\hat{\beta} - r)' \left[R\,\mathrm{Var}(\hat{\beta})R'\right]^{-1} (R\hat{\beta} - r) \sim \chi_q^2$$

选择卡方统计量而不是 t 统计量或者正态分布统计量的原因是卡方分布能检验多个参数。实际上如果只有一个参数，那么检验就是回归中的 t 检验。小样本时，多个系数同时为 0 的 F 检验也是 wald 检验的一种形式（小样本方差不准确，有自由度，大样本基于大数定理趋于期望值，整体变成卡方分布）。

$$F = \frac{W/q}{SSR/(n-k)} \sim F_{q,n-k}$$

注[渐进性的检验]

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{d} Normal(0, \boldsymbol{V})$$

V 是 $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$ 的渐近方差(随机变量的方差在样本量趋近于无穷时的极限值),可记为 $Avar\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) = V$

Wald 检验：

若 $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{d} Normal(0, \boldsymbol{V})$, 对于任意非随机矩阵, 有 $R_{Q\times P}, Rank(R) = Q$,

$$\sqrt{N}\boldsymbol{R}(\widehat{\boldsymbol{\theta}}_N-\boldsymbol{\theta}) \xrightarrow{d} Normal(0,\boldsymbol{RVR}')$$
$$\left[\sqrt{N}\boldsymbol{R}(\widehat{\boldsymbol{\theta}}_N-\boldsymbol{\theta})\right]'[\boldsymbol{RVR}']^{-\mathbf{1}}[\sqrt{N}\boldsymbol{R}(\widehat{\boldsymbol{\theta}}_N-\boldsymbol{\theta})] \xrightarrow{d} \chi^2_Q$$

Delta method:

$$\sqrt{N}[\boldsymbol{c}(\widehat{\boldsymbol{\theta}}_N)-\boldsymbol{c}(\boldsymbol{\theta})] \xrightarrow{d} Normal(0,\boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{V}\boldsymbol{C}(\boldsymbol{\theta})')$$
$$\left\{\sqrt{N}[\boldsymbol{c}(\widehat{\boldsymbol{\theta}}_N)-\boldsymbol{c}(\boldsymbol{\theta})]\right\}'[\boldsymbol{RVR}']^{-1}\{\sqrt{N}[\boldsymbol{c}(\widehat{\boldsymbol{\theta}}_N)-\boldsymbol{c}(\boldsymbol{\theta})]\} \xrightarrow{d} \chi^2_Q$$

定义 $C(\theta)=\nabla_\theta c(\theta)$ 为函数的维雅可比矩阵 (Jacobian)(每一行使用相同的函数, 每一列使用相同的变量, 求偏导)。

例题推荐

在多重假设下的 wald 检验量:

$\left[\sqrt{n}\left(\mathbf{R}\hat{\beta}-r\right)\right](\sigma^2\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}\left[\sqrt{n}\left(\mathbf{R}\hat{\beta}-\mathbf{r}\right)\right] \overset{a}{\sim} \chi^2_q$

$W=(\mathbf{R}\hat{\beta}-\mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta}-\mathbf{r})/\hat{\sigma}^2$

异方差存在时，只需要修改形式:

$$\widehat{\mathrm{AVar}(\hat{\boldsymbol{\beta}})}=[n/(n-k-1)]\,\mathbf{X}'\mathbf{X}]^{-1}\left(\sum_{t=1}^{n}\hat{u}_t^2\mathbf{x}_t'\mathbf{x}_t\right)(\mathbf{X}'\mathbf{X})^{-1}$$

stata 中的 robust 检验使用的是 wald 统计量而不是 F 统计量。

但也可以转化为 F 统计量

$$\frac{(\hat{R}\hat{\beta}-q)'[\hat{R}(X'X)^{-1}\hat{R}']^{-1}(\hat{R}\hat{\beta}-q)/J}{s^2}=\frac{(e'e-e'e)/J}{e'e/(n-K)}$$

第五章 工具变量

IV 思想：将内生变量的内生部分排除
IV 估计思想：矩估计
工具变量的选择

5.1 内生性与外生性

计量经济学导论 15 章内生性导致了什么后果

定义 5.1 (外生性、内生变量)

概念理解。
弱外生性：$cov(u,x)=0$。
强外生性：$cov(u,f(x))=0$，或者写成 $E(u_i|x)=0$。随机扰动项与解释变量的任何函数都不相关，即零条件均值假定 (MLR.4) 成立
具有内生性，也就是至少连弱外生性也没有满足 $cov(u,x)\neq 0$ ♣

好的控制变量——前置的变量
方法：DID RDD DDD FIX IV Bunching
偏相关和自相关

定义 5.2 (内生变量概念理解)

内生：模型内部其他经济变量决定
外生：模型外部其他因素决定

$$y=\beta x+u$$

如果 **x** 和误差项相关，则是内生的；不相关，则是外生的。在因果分析中，模型变量为解释变量部分（也就是等号右侧，包含误差项）
误差：系统误差和抽样误差（样本统计量上的误差，样本的数字特征，例如样本方差等，样本）
如果一个方差同时有外生变量和内生变量。♣

内生性的原因可能来自：

- 遗漏变量 (omitted variables)
- 测量误差 (measurement error)
- 联立相关 (simultaneity)

5.1.1 控制变量内生

注[重点理解] 如果解释变量外生，控制变量内生，也有**可能**使得估计有偏。注意！是可能，可能。

$$y_i=\beta_0+\beta_1 x_{i1}+\beta_{i2}+u_i$$

此时假设 $E(u_i|x_1)=0, cov(u_i,x_2)\neq 0 \longrightarrow E(u_i|x_1,x_2)\neq 0$。
通过 fwl3.1定理求得系数

$$\hat{\beta}_1 = \frac{\sum \hat{\gamma}_{i1} y}{\sum \hat{\gamma}_{i1}^2}, \hat{\beta}_2 = \frac{\sum \hat{\gamma}_{i2} y}{\sum \hat{\gamma}_{i2}^2}$$

接下来考察 $x_1 x_2$ 之间的回归关系，也就是上面的系数。

$$\hat{\gamma_{i1}} = x_{i1} - \hat{\delta}_0 - \hat{\delta}_1 x_{i2}, \hat{\gamma_{i2}} = x_{i2} - \hat{\delta}_0 - \hat{\delta}_1 x_{i1}$$

此时内生性的控制变量系数一定是有偏的；而外生性系数是否有偏取决于与内生变量的相关性。内生变量和 x_1 无关则无偏，相关则有偏。

$$E\left(\hat{\beta}_1\right) = \beta_1 + \frac{\sum \hat{\gamma}_{i1} E\left(u_i | x_1, x_2\right)}{\sum \hat{\gamma}_{i1}^2} \overset{?}{=} \beta_1$$

大样本情况下求极限：

$$\begin{aligned}
\hat{\beta}_2 &= \beta_2 + \frac{\sum (x_{i2} - \hat{\sigma}_0 - \hat{\sigma}_1 x_{i1}) u_i}{\sum \hat{\gamma}_{i2}^2} \longrightarrow lim(\hat{\beta}_2) \\
&= \beta_2 + \frac{E(x_{i2} u_i) - \hat{\sigma}_0 E(u_i) - \hat{\sigma}_1 E(x_{i1} u_i)}{E(\hat{\gamma}_{i2}^2)} = \beta_2 + \frac{E(x_{i2} u_i)}{E(\hat{\gamma}_{i2}^2)} \neq \beta_2 \\
\hat{\beta}_1 &= \beta_1 + \frac{\sum (x_{i1} - \hat{\delta}_0 - \hat{\delta}_1 x_{i2}) u_i}{\sum \hat{\gamma}_{i1}^2} \longrightarrow lim(\hat{\beta}_1) \\
&= \beta_1 + \frac{E(x_{i1} u_i) - \hat{\delta}_0 E(u_i) - \hat{\delta}_1 E(x_{i2} u_i)}{E(\hat{\gamma}_{i1}^2)} = \beta_1 - \hat{\delta}_1 \frac{E(x_{i2} u_i)}{E(\hat{\gamma}_{i1}^2)} \overset{?}{=} \beta_1
\end{aligned}$$

5.1.2 控制无关变量

真实世界 a 式子：

$$y = \beta_1 x_1 + u$$

我们选择了 b 式子

$$y = \beta_1 x_1 + \beta_2 x_2 + u$$

假设都是外生变量 $E(u|x_1, x_2) = 0$

$$\hat{\beta}_{b1} = \beta_1 + \beta_2 (x_1' M x_1)^{-1} (x_1' M u)$$

其方差为 $var(\hat{\beta}_{b1}|x) = \sigma^2 (x_1' M x_1)^{-1}$
而式子 a 的方差为 $var(\hat{\beta}_{a1}|x) = \sigma^2 (x_1' x_1)^{-1}$
$var(\hat{\beta}_{b1}|x) - var(\hat{\beta}_{a1}|x)$ 是一个**半正定矩阵**。
而这种效率的损失其实就可以表现为前面的**多重共线性**的情况。

5.1.3 遗漏变量

先写方程形式：
假设正确模型为：

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

遗漏以后：

$$y_i = \beta_0 + \beta_1 x_{i1} + v_i$$

带入估计量：

$$\tilde{\beta}_1 = \frac{\sum(x_{i1} - \bar{x}_1)(y_i - \bar{y})}{\sum(x_{i1} - \bar{x}_1)^2}$$

$$\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x}_1$$

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2}$$

后半部分就是 x_2 对 x_1 进行回归。

$$E(\tilde{\beta}_0) = \beta_0 + \beta_2 \left[\bar{x}_2 - \bar{x}_1 \frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2}\right]$$

遗漏变量如果和 x_1 相关，则 $\beta_0\beta_1$ 均有偏误，如果不相关，则 β_1 无偏，但则 β_0 有偏。

扩展到一般情况，假设有 k 个变量，遗漏了第 k 个变量。

$$E(\tilde{\beta}_j) = \frac{\sum \hat{\gamma}_{ij} y}{\sum \hat{\gamma}_{ij}^2} = \beta_j + \beta_k \boxed{\frac{\sum \hat{\gamma}_{ij} x_{ik}}{\sum \hat{\gamma}_{ij}^2}}$$

方框式子利用了 fwl3.1定理。$\hat{\gamma}_{ij}$ 代表 x_j 对 $x_1 - x_{k-1}$ 回归的残差。因为遗漏了 x_k，所有未能消掉 y 中包含的 $\beta_k x_k$。

在遗漏了 x_2 的情况下——接下来用矩阵形式表达，其实就是：

$$\hat{\beta}_1 = \beta_1 + \beta_2 (x_1' M x_1)^{-1}(x_1' M x_2) + \beta_2 (x_1' M x_1)^{-1}(x_1' M u)$$

遗漏变量导致的偏误方向取决于两个关系：

y 和 x_2 的方向；x_2 和 x_1 的方向。

	$Partial\ Corr(x_j, x_k) > 0$	$Partial\ Corr(x_j, x_k) < 0$
$\beta_k > 0$	偏差为正	偏差为负
$\beta_k < 0$	偏差为负	偏差为正

在计量实践中，我们偏好于低估效应————这样实际情况比我们研究的结果还要大，能作证明研究的有效性。

例题 5.1 推荐参考：政府一次性投资与产业长期发展。这篇 AER 的文章就交代自己估计的研发乘数效应偏低了（估计的是 0.3），其他文章估的数值是 0.6。因为他遗漏了大学的科研变量。

遗漏相关变量后方差为

$$Var(\tilde{\beta}_1) = \frac{\sigma_v^2}{\sum(x_{i1} - \bar{x}_1)^2} \approx \frac{\sigma_u^2 + \beta_2^2 \mathrm{Var}(x_2)}{\sum(x_{i1} - \bar{x}_1)^2}$$

这里使用约等于是因为计算方差时，分母自由度其实是不同的，遗漏变量是 n-2,，没有遗漏是 n-1。

$$E(\hat{\sigma}_v^2) = E\left(\frac{SSR_v}{n-2}\right) \neq E\left(\frac{SSR_u}{n-3}\right) = \sigma_u^2$$

而实际方差为

$$Var\left(\hat{\beta}_1\right)=\frac{\sigma_u^2}{SST_1\left(1-R_1^2\right)}=\frac{\sigma_u^2}{\sum\left(x_{i1}-\bar{x}_1\right)^2\left(1-R_1^2\right)}$$

两者大小不一定谁更大或者小。

5.1.4 测量误差

定义 5.3 (测量误差)

测量误差是指变量的实际观测值与真实值存在差异。测量误差可以分为被解释变量测量误差和解释变量测量误差。测量误差的来源有两类: 一类是登记性误差; 二是被调查对象故意瞒报真实值, 比如, 明星谎报收入。 ♣

假设真实模型应该是

$$y_i^*=\beta_0+\beta_1x_i+u_i$$

但是 y^* 难以直接观察，只能通过测量得到

$$y_i=y_i^*+\varepsilon_i$$

$$y_i=\beta_0+\beta_1x_i+v_i=\beta_0+\beta_1x_i+u_i+\varepsilon_i$$

只要这部分测量误差和 x 无关 $coc(\varepsilon,x)$，估计依旧无偏。

y_i 存在测量误差	OLS 估计量	OLS 估计量方差
$E(\varepsilon_i\|x_i)=0$	无偏	比无测量误差时大
$E(\varepsilon_i\|x_i)\neq 0$	有偏	比无测量误差时大

但是无论是否相关，方差都会被放大。
简化假设 $\varepsilon_i\perp u_i$(但是这个假设一定满足吗？)，得到

$$Var(v_i)=Var(u_i+\varepsilon_i)=Var(u_i)+Var(\varepsilon_i)+2Cov(u_i,\varepsilon_i)$$

$$\operatorname{Var}(\widehat{\beta}_1\mid X)=\frac{\sigma_u^2+\sigma_\varepsilon^2}{S_{xx}}>\frac{\sigma_u^2}{S_{xx}}$$

✍ **练习 5.1** 在教育回报问题当中

$$log(wage)=\beta_0+\beta_1educ+u$$

假设工资存在测量误差，$E(\varepsilon_i|educ)=0$ 是否成立？

个人想法，教育程度越高越容易瞒报（灰色收入，疑心重，隐私性强 $u\leq 0$），那么系数会低估 $\hat{\beta}=\beta+\frac{\sum ux_i}{\sum x_i^2}$; 如果教育程度越高越想要炫耀，可能会高估。

比较棘手的是解释变量的测量误差。

$$y_i=\beta_0+\beta_1x_i^*+u_i$$

$$x_i = x_i^* + \varepsilon_i$$

$$y_i = \beta_0 + \beta_1 x_i + v_i = \beta_0 + \beta_1 x_i + u_i - \beta_1 \varepsilon_i$$

此时无偏性条件为 $cov(u_i - \beta_1 \varepsilon_i, x_i) = 0$, 但是 $x_i = x_i^* + \varepsilon_i$, 此时 $cov(u_i - \beta_1 \varepsilon_i, x_i) = cov(u_i - \beta_1 \varepsilon_i, x_i^* + \varepsilon_i)$。最后 ε_i 自身和自身相关了。例如幸存者偏差，对幸存者进行测量，未观察到的干扰项同时影响了测量结果和回归结果。例如只在健身馆调研锻炼的作用。

$$\begin{aligned}
\tilde{\beta}_1 &= \frac{\sum (x_{i1} - x_1)(y_i - y)}{\sum (x_{i1} - \bar{x}_1)^2} \\
&= \beta_1 + \frac{\sum (x_{i1} - \bar{x}_1) u_i}{\sum (x_{i1} - \bar{x}_1)^2} - \beta_1 \frac{\sum (x_{i1} - \bar{x}_1) \varepsilon_i}{\sum (x_{i1} - \bar{x}_1)^2} \\
&= \beta_1 \left[\frac{\sum (x_{i1} - \bar{x}_1)^2 - \sum (x_{i1} - \bar{x}_1) \varepsilon_i}{\sum (x_{i1} - \bar{x}_1)^2} \right] + \frac{\sum (x_{i1} - \bar{x}_1) u_i}{\sum (x_{i1} - \bar{x}_1)^2}
\end{aligned}$$

使用大数定理求概率极限

$$\begin{aligned}
plim(\tilde{\beta}_j) &= \beta_1 plim \left[\frac{\sum (x_{i1} - \bar{x}_1)^2 - \sum (x_{i1} - \bar{x}_1) \varepsilon_i}{\sum (x_{i1} - \bar{x}_1)^2} \right] \\
&= \beta_1 \left[\frac{\sigma_x^2 - Cov(x_1, \varepsilon_i)}{\sigma_x^2} \right]
\end{aligned}$$

使用经典误差假设 $Cov(x_i^*, \varepsilon_i) = 0$（真实值和误差值不相关，不同测量中产生的误差也不相关）。

$$Cov(x_1, \varepsilon_i) = Cov\left(x_i^* + \varepsilon_i, \varepsilon_i\right) = V = \sigma_\varepsilon^2$$

$$plim(\tilde{\beta}_j) = \beta_1 \left[\frac{\sigma_x^2 - \sigma_\varepsilon^2}{\sigma_x^2} \right] < \beta_1$$

在经典测量误差假定下, 解释变量测量误差会导致估计向下偏误。

更直觉的理解是测量偏误稀释了 x 的信号，散点图变得更加离散，最终会倾向于拉平回归线。

练习 5.2 解释变量测量误差会导致估计量方差增大

证明

$$\mathrm{Var}(\tilde{\beta}_1 | \mathbf{X}) = \frac{\mathrm{Var}(v_i | \mathbf{X})}{\sum (x_i - \bar{x})^2}$$

$$\mathrm{Var}(v_i) = \mathrm{Var}(u_i - \beta_1 \varepsilon_i) = \mathrm{Var}(u_i) + \beta_1^2 \mathrm{Var}(\varepsilon_i) - 2\beta_1 \mathrm{Cov}(u_i, \varepsilon_i)$$

经典测量误差假定 $\mathrm{Cov}(u_i, \varepsilon_i) = 0$:

$$\mathrm{Var}(v_i) = \sigma_u^2 + \beta_1^2 \sigma_\varepsilon^2$$

5.1.5 联立相关

定义 5.4 (联立相关)

联立相关是指解释变量和被解释变量在同一时期存在相互影响。
y_1, y_2 相互影响，那么可以写为：

$$\begin{cases} y_{i1} = \beta_0 + \beta_1 y_{i2} + \alpha_2 x_{i1} + \cdots + \alpha_k x_{ik} + u_i \\ y_{i2} = \gamma_0 + \gamma_1 y_{i1} + \alpha_2 z_{i1} + \cdots + \alpha_k z_{ik} + v_i \end{cases}$$

♣

✍ **练习 5.3 证明内生性** x 和 z 都为外生变量。

证明 [内生性]

$$\mathrm{Cov}(a+b, c) = \mathrm{Cov}(a, c) + \mathrm{Cov}(b, c)$$

$$Cov(y_2, u_i) = Cov(\gamma_1 y_1 + v_i, u_i) = Cov(\gamma_1 y_1, u_i)$$
$$= Cov(\gamma_1 \beta_1 y_2 + \gamma_1 u_i, u_i) = \gamma_1 \beta_1 Cov(y_2, u_i) + \gamma_1 Var(u_i)$$

$$Cov(y_2, u_i) = \frac{\gamma_1 Var(u_i)}{1 - \gamma_1 \beta_1} \neq 0$$

例题 5.2 联立方程

更加直观的例子。

再比如求解市场均衡供求数量与价格的均衡：

$$\begin{cases} q_i^d = \alpha_0 + \alpha_1 p_i + u_i \\ q_i^s = \beta_0 + \beta_1 p_i + v_i \end{cases}$$

最后会解得市场价格是误差项的函数 $p_i = f(u_i, v_i) = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1}$。在式子中，这个就表现为违反了内生性。

✍ **练习 5.4 联立偏误** 联立使得 $\hat{\beta}_1, \hat{\gamma}_1$ 有什么偏误？

证明 [联立偏误]

$$E[\widehat{\beta}_1 \mid X] = \beta_1 + E\left[\frac{\sum_i \widehat{\delta}_{i1} u_i}{\sum_i \widehat{\delta}_{i1}^2} | X\right]$$

$\frac{\sum_i \widehat{\delta}_{i1} u_i}{\sum_i \widehat{\delta}_{i1}^2}$ 也是一个 fwl 组成的回归。
使用 u_i 对 $\widehat{\delta}_{i1}$ 回归，而

$$\widehat{\delta}_{i1} = y_{i2} - \sum \delta_{il} x_{il} = \gamma_0 + \gamma_1 y_{i1} + \alpha_2 z_{i1} + \cdots + \alpha_k z_{ik} + v_i - \sum \delta_{il} x_{il} - \beta_0$$

$$u_i = y_i 1 =$$

能进一步简化吗

5.1.6 控制变量的选择

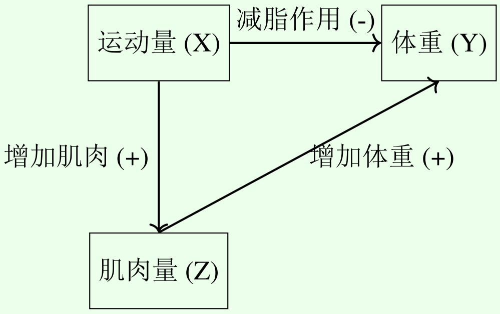
相关、偏相关、独立、外生。

独立是最强的性质，如果变量之间彼此独立，那么自相关和偏相关一定为 0。

偏相关、相关衡量的都是线性关系。

相关 =0，偏相关不一定 =0。xy 的关系可能被 z 掩盖抑制了（抑制、遮蔽效应），例如 z 和 x 正影响，z 和 y 负影响。

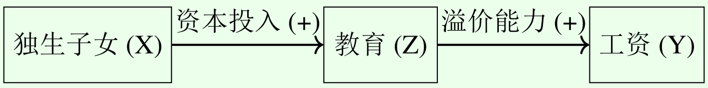
其逆否命题例子一样——偏相关 $\neq 0$，相关不一定 $\neq 0$。



直观上，相关关系包含了直接和间接影响，而直接和间接内部存在加减，偏相关会去除某些间接部分。

相关 $\neq 0$，偏相不一定 $\neq 0$。x 通过 z 影响 y，此时相关，控制了 z 后不再相关。

其逆否命题例子一样——偏相关 =0，相关不一定 =0。



这个例子参考论文《Little emperors in the workplace: Labor market consequences of China's one-child policy》。著名的录取歧视和中介谬误也是这个原理，参见论文《Sex Bias in Graduate Admissions: Data from Berkeley》

相关性和偏相关参考资料

命题 5.1

变量独立，则相关系数一定为 0。

证明 [独立性和相关性]

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$X \perp Y \implies \mathbb{E}[XY] = \mathbb{E}[X]\cdot\mathbb{E}[Y]$$

同理，独立，也必然外生。

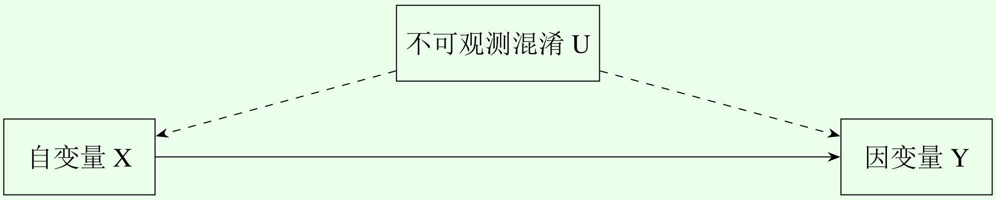
命题 5.2

外生性满足，一定不线性相关。就是 0 条件均值保持无偏性。

但是外生性不能保证偏相关为 0。

以上情况告诉了我们，**控制变量要控制前置变量**。这里使用《the book of why》的框架。

前门路径: X → M → Y, U 不可观测



后门路径: X ← U → Y

后门的典型例子就是冰淇淋销售量和鲨鱼袭击人的数量，数据呈现相关性是因为他们都容易在夏天发生，夏天就是那个造成干扰的 U。

前门的意思则是不能随意控制。

选择前置的控制变量就是不要随意控制 x 机制上的中介，但要控制会同时影响 x 和 y 的变量。

5.1.7 代理变量

真实模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u$$

此时 x_3^* 无法直接观测，例如能力、IQ... 使用代理变量：

$$x_3^* = \delta_0 + \delta_1 x_3 + v$$

带入可得

$$y = \beta_0 + \delta_0\beta_3 + \beta_1 x_1 + \beta_2 x_2 + \beta_3\delta_1 x_3 + u + \beta_3 v$$

$$y = \beta_0^* + \beta_1 x_1 + \beta_2 x_2 + \beta_3^* x_3 + u^*$$

β_1, β_2 无偏估计要求：

1. $E(u|x_1, x_2, x_3^*) = 0.$
   由于真实模型满足经典线性模型假定，这一假定自然满足（即 x_1, x_2, x_3^* 均为外生变量）。
2. $E(u|x_3) = 0.$
   x_3 是 x_3^* 的代理变量。此假设要求代理变量 x_3 本身与真实扰动项 u 不相关。若 $E(u|x_3^*) = 0$，则 $E(u|x_3) = 0$ 成立，这要求 x_3 仅通过 x_3^* 影响 y。
3. $E(v|x_1, x_2, x_3) = 0.$
   要求 x_3 是 x_3^* 的一个好的代理变量。
   （更严格的代理变量条件是 x_3^* 只与 x_3 线性相关，且 x_3 与其他解释变量无关。）

5.2 工具变量

工具变量是矩估计的思想。

隐含参数的总体用矩阵表示，然后用样本矩阵去表示。其中 $E(z'u)$

命题 5.3 (iv 假设)

iv 有两个假设

假设 1：$cov(x,z)\neq 0$。这个假定限制了 z 会影响 z。

假定 2：$cov(u,z)\neq 0$。这个假定限制了 z 只通过 x 影响 y，排除了其他影响路径。排他性假设是无法被检验的，1 是无法排除所有路径，2 是两阶段的检验也无法无偏检验。

练习 5.5 排他性为什么不能使用以下检验 z 是 iv，x 是解释变量。

$$y=\beta_0+\beta_1 z+v$$

$$y=\gamma_0+\gamma_1 x_1+\gamma_2 z+u$$

证明 [为什么不能通过以上来检验]

对于第一个方程，如果 z 会影响 x 同时 z 会影响 y，

那么第二个方程也是内生有偏的。u 同时会影响 x, 同时 z 和 x 又相关。

核心原因就在于 x 本身的内生性无法通过控制 z 来解决或者以此判断 z 的可靠性。

两个条件本身存在内在矛盾，x 和 z 越相关，z 和 y 往往又越来越直接相关。

寻找 iv：同伴效应、关系之关系、遥远的历史关系、天文地理。

$$Cov(z,y)=\beta_1 Cov(z,x)+Cov(z,u)$$

$$\beta_1=\frac{Cov(z,y)}{Cov(z,x)}=\frac{Cov(z,y)/Var(z)}{Cov(z,x)/Var(z)}$$

$\frac{Cov(z,y)/Var(z)}{Cov(z,x)/Var(z)}$ 其实就是 reg y z 系数与 reg x z 系数只比。

$$\hat{\beta}_1=\frac{\sum(z_i-\bar{z})(y_i-\bar{y})}{\sum(z_i-\bar{z})(x_i-\bar{x})}=\beta_1+\frac{\sum(z_i-\bar{z})u_i}{\sum(z_i-\bar{z})(x_i-\bar{x})}$$

$$\widehat{\beta}_{\mathrm{IV}}=(z'x)^{-1}z'y$$

5.2.1 期望

注[条件期望 iv] iv 的条件期望相对 z 是无偏的，相对 z,x 是有偏的。因此实际上 iv 估计量**并不无偏**，重要的是**一致性**。

例如小样本情况下，iv 估计量是有偏的：

$$E\left(\hat{\beta}_1\right)=\beta_1+\frac{\sum(z_i-z)E(u_i|x_i,z_i)}{\sum(z_i-\bar{z})(x_i-\bar{x})}\neq\beta_1$$

大样本性质：

$$plim(\hat{\beta}_1)=\beta_1+\frac{Cov(z,u)}{Cov(z,x)}=\beta_1$$

注[表述] 模型指的是估计方程的形式。iv 和 ols 是不同的估计方法而不是不同的模型。

练习 5.6 多元回归求解

$$y_1=\beta_0+\beta_1 y_2+\beta_2 z_1+\cdots+\beta_k z_{k-1}+u$$

y_2 内生，z_k 工具变量。
多元线性情况相同：

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u$$

y_2 内生，z_k 工具变量。
得到

$$\begin{aligned} Cov(z_k, u) &= Cov(z_k, y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_2 - \ldots - \beta_k z_{k-1}) \\ &\longrightarrow \sum z_k (y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_2 - \ldots - \beta_k z_{k-1}) = 0 \\ &\longrightarrow \sum z_k y_1 - \beta_1 \sum z_k y_2 = 0 \end{aligned}$$

$$\widehat{\beta}_1 = \frac{Cov(z_k, y_1)}{Cov(z_k, y_2)} = \beta_1 + \frac{Cov(z_k, u)}{Cov(z_k, y_2)}$$

5.2.2 方差

$$\hat{\beta}_1 = \frac{\sum (z_i - \bar{z})(y_i - \bar{y})}{\sum (z_i - \bar{z})(x_i - \bar{x})} = \beta_1 + \frac{\sum (z_i - \bar{z}) u_i}{\sum (z_i - \bar{z})(x_i - \bar{x})}$$

工具变量的条件方差以 z 为条件。
渐进情况下 $\sqrt{N}(\hat{\beta} - \beta)$ 遵循 $N^{-1/2} \sum_{i=1}^{\hat{N}} z_i' u_i$ 的正态分布。
此时引出**假设 3**：

$$E(u^2 z' z) = \sigma^2 E(z' z)$$

这个假设等价于 $E(u^2|z) = \sigma^2$
因此得到渐进方差：$Avar[\hat{\beta}^{2sls} - \beta] = \sigma^2 \{E(x'z)[E(z'z)]^{-1} E(z'x)\}^{-1}$
方差估计为 $\hat{\sigma}^2 \equiv (N-K)^{-1} \sum_{i=1}^{N} \hat{u}_i^2$

$$[Cov(z, x)]^2 = (\rho_{z,x} \cdot \sigma_z \cdot \sigma_x)^2 = \rho_{z,x}^2 \cdot \sigma_z^2 \cdot \sigma_x^2$$

渐进方差为：

$$plim\left[(\hat{\beta}_1 - \beta_1)^2\right] = \frac{\sigma^2}{n \sigma_x^2 \rho_{z,x}^2}$$

对于 **K-K 的矩阵**来说，可以得到 $\hat{\sigma}^2 \left(\sum_{i=1}^{N} \hat{x}_i' \hat{x}_i\right)^{-1} = \hat{\sigma}^2 (\hat{X}' \hat{X})^{-1}$

命题 5.4 (iv 方差)

一阶段相关性越弱，方差越大。

注

考过的证明：

$$Avar \hat{\beta}^{2sls} - Avar \hat{\beta}^{ols} > 0$$

教材 5.2.3 97 页，转化为结果是否正定或者半正定

书中的例子是比较 K-K 工具变量的 2sls 和其他 L-K 的工具变量 (工具变量数大于内生变量数) 的估计。

引入 L-K 的非随机矩阵 Γ，使用其作为工具变量，得到线性估计 $\tilde{\mathbf{x}} \equiv \mathbf{z}\Gamma$

此时 2SLS 的估计工具变量式子为：$\mathbf{x}^* = \mathbf{z}\Pi, \Pi = [\mathbf{E}(\mathbf{z}'\mathbf{z})]^{-1}\mathbf{E}(\mathbf{z}'\mathbf{x}) \equiv \mathbf{D}^{-1}\mathbf{C}$

代入 2sls 的样本渐进方差：

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sigma^2[\mathrm{E}(\mathbf{x}^{*\prime}\mathbf{x}^*)]^{-1}$$

同理，其他 iv 估计的样本渐进方差为 (代入渐进方差矩阵估计式子即可)：

$$\mathrm{Avar}[\sqrt{N}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})] = \sigma^2[\mathrm{E}(\tilde{\mathbf{x}}'\mathbf{x})]^{-1}[\mathrm{E}(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})][\mathrm{E}(\mathbf{x}'\tilde{\mathbf{x}})]^{-1}$$

由于 $x = x^* + r$(x^* 是工具变量的拟合值)

因此 $E(\tilde{x}'x) = E(\tilde{x}'(x^* + r)) = E(\tilde{x}'x^*) + E(\tilde{x}'r)$

工具变量估计中 $x = \beta z + r = x^* + r$ 也就是 $r = x - x^* = r - \beta z$，因此 $E(z'r) = 0$ 。

进一步可以看出 $E(x^*r) = E(\beta zr) = 0$，同理，L-K 的工具变量可以得到 $E(\tilde{x}r) = 0$

得到 $E(\tilde{x}r) = 0$ 后，代入 $E(\tilde{x}'x) = E(\tilde{x}'(x^* + r)) = E(\tilde{x}'x^*) + E(\tilde{x}'r)$ 即可得到 $\mathbf{E}(\tilde{\mathbf{x}}'\mathbf{x}) = \mathbf{E}(\tilde{\mathbf{x}}'\mathbf{x}^*)$

在后文比较有效性时即可便于合并同类项然后构造对称矩阵，进而发现是 (半) 正定。

接下来比较 K-K 工具变量组合和 L-K 工具变量组合的有效性：

$$\begin{aligned}
&\mathrm{Avar}[\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] - \mathrm{Avar}[\sqrt{N}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\
&= \sigma^2\left\{\mathrm{E}(\mathbf{x}^{*\prime}\mathbf{x}^*) - [\mathrm{E}(\tilde{\mathbf{x}}'\mathbf{x})^{-1}\mathrm{E}(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})\mathrm{E}(\mathbf{x}'\tilde{\mathbf{x}})]^{-1}\right\}
\end{aligned}$$

因为重点是比较正负性来判断有效性，因此只需要关注大括号以内的部分，展开即可得到下面式子

$$\begin{aligned}
&\mathrm{E}(\mathbf{x}^{*\prime}\mathbf{x}^*) - \mathrm{E}(\mathbf{x}'\tilde{\mathbf{x}})[\mathrm{E}(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})]^{-1}\mathrm{E}(\tilde{\mathbf{x}}'\mathbf{x}) \\
&= \mathrm{E}(\mathbf{x}^{*\prime}\mathbf{x}^*) - \mathrm{E}(\mathbf{x}^{*\prime}\tilde{\mathbf{x}})[\mathrm{E}(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})]^{-1}\mathrm{E}(\tilde{\mathbf{x}}'\mathbf{x}^*) = \mathrm{E}(\mathbf{s}^{*\prime}\mathbf{s}^*)
\end{aligned}$$

最后一步的补充：

看作 x^* 对 x 回归

$s^* = x^* - \tilde{x}\phi = x^* - \tilde{x}E(\tilde{x}'\tilde{x})^{-1}E(\tilde{x}'x^*)$

5.2.3 拟合优度

拟合优度计算式子为

$$R^2 = 1 - \frac{SSR}{SST} < 0$$

对于 ols 来说：

$$SSR = y'(I - P_X)'(I - P_X)y = y'(I - P_X)y \geq 0$$

但是 iv 是使用 z 去预测 x。

$$\hat{y} = X\hat{\beta}_{IV} = X(Z'X)^{-1}Z'y = P_{XZ}y$$

此时投影不一定是正交的。

$$R^2 = 1 - \frac{(y - X\hat{\beta}_{IV})'(y - X\hat{\beta}_{IV})}{(y - \bar{y})'(y - \bar{y})}$$

这时候有可能出现负拟合优度的情况。但实际上 iv 的拟合优度没有特别意义

5.3 2SLS

具体使用**矩估计**的方法表达：
以最简单的情况为例子：工具变量和残差不相关，保证外生性，和 x 相关。

$$\begin{cases} E(z'u) = 0 \\ cov(z, x) \neq 0 \\ rank(z'x) = k \end{cases}$$

对于方程：

$$y = \beta_0 + \beta_1 x$$

在方程两侧同时乘以 z'
方阵满列满秩时得到唯一解

$$rank\ E(z'x) = K$$

rank condition: 秩条件

$$\text{(a) rank E(z'z) = L; (b) rank E(z'x) = K.}$$

K 代表 K 个内生解释变量，L 代表 L 个工具变量。

引理 5.1 (秩条件)

$$\text{(a) rank E(z'z) = L; (b) rank E(z'x) = K.}$$

这说明多少个内生变量就至少多少个工具变量。这个是为了保证方程具有唯一解。
也就是每个内生变量至少有一个相关的外生 iv。

$$Z'X = \begin{pmatrix} \mathrm{Cov}(z_1, x_1) & \dots & \mathrm{Cov}(z_1, x_K) \\ \vdots & & \vdots \\ \mathrm{Cov}(z_L, x_1) & \dots & \mathrm{Cov}(z_L, x_K) \end{pmatrix}$$

秩条件用于判断 iv 是否足够。

命题 5.5 (单一工具变量，iv 和 2sls 等价)
单一工具变量，iv 和 2sls 等价

证明 [单一工具变量，iv 和 2sls 等价] 证明如下
方程形式：
一阶段回归

$$X_i = \gamma_0 + \gamma_1 Z_i + v_i$$

$$\hat{X}_i - \bar{X} = \hat{\gamma}_1 (Z_i - \bar{Z})$$

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Z_i - \bar{Z})}{\sum_{i=1}^{n}(Z_i - \bar{Z})^2}$$

对于估计式子：

$$Y_i = \beta_0^* + \hat{\beta}_{2SLS}\hat{X}_i + \epsilon_i$$

$$\hat{\beta}_{2SLS} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(\hat{X}_i - \bar{X})}{\sum_{i=1}^{n}(\hat{X}_i - \bar{X})^2}$$

此时代入有：

$$\hat{\beta}_{2SLS} = \frac{\hat{\gamma}_1 \sum(Y_i - \bar{Y})(Z_i - \bar{Z})}{(\hat{\gamma}_1)^2 \sum(Z_i - \bar{Z})^2} = \frac{\sum(Y_i - \bar{Y})(Z_i - \bar{Z})}{\hat{\gamma}_1 \sum(Z_i - \bar{Z})^2}$$

再代入 $\hat{\gamma}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Z_i - \bar{Z})}{\sum_{i=1}^{n}(Z_i - \bar{Z})^2}$

$$\hat{\beta}_{2SLS} = \frac{\sum(Y_i - \bar{Y})(Z_i - \bar{Z})}{\left[\frac{\sum(X_i - \bar{X})(Z_i - \bar{Z})}{\sum(Z_i - \bar{Z})^2}\right]\sum(Z_i - \bar{Z})^2} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(Z_i - \bar{Z})}{\sum_{i=1}^{n}(X_i - \bar{X})(Z_i - \bar{Z})}$$

多元时：

全部换成残差形式：

外生控制变量为集合 w。

$\dot{z}$ 为 z 对控制变量回归后的残差项。

$\dot{x}$ 为 x 对控制变量回归后的残差项。

首先得到式子

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n}\dot{x}\dot{z}}{\sum_{i=1}^{n}\dot{z}^2}$$

然后得到估计式子

$$\hat{\beta}_{2sls} = \frac{\sum_{i=1}^{n}\dot{\hat{x}}y}{\sum_{i=1}^{n}\dot{\hat{x}}^2}$$

这时候转化

$$\dot{\hat{x}} = \hat{\gamma}_1\dot{z}$$

$$\hat{\beta}_{2sls} = \frac{\sum_{i=1}^{n}\dot{\hat{x}}y}{\sum_{i=1}^{n}\dot{\hat{x}}^2} = \frac{\sum_{i=1}^{n}\dot{z}y}{\hat{\gamma}_1\sum_{i=1}^{n}\dot{z}^2} = \frac{\sum_{i=1}^{n}\dot{z}y}{\sum_{i=1}^{n}\dot{z}\dot{x}}$$

矩阵形式：

$$P_Z = Z(Z'Z)^{-1}Z$$

$$\hat{X} = P_Z X = Z(Z'Z)^{-1}ZX$$

$$\begin{aligned}\beta_{2sls} &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y\\ &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'Y\end{aligned}$$

在 L=K 的情况下，式子才可以进一步化简

$$\begin{aligned}\beta_{2sls} &= (\hat{X}'\hat{X})^{-1}\hat{X}'Y\\ &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'Y\\ &= (Z'X)^{-1}X'Y\end{aligned}$$

也就是

$$\beta = [E(z'x)]^{-1}E(z'y) = [E(zz')^{-1}E(z'x)]^{-1}[E(zz')^{-1}E(z'y)]$$

也就是 y 对 z 回归，减去 x 对 z 回归的部分。此时如果 $cov(z,x)=0$，则无法取逆矩阵，自然无法求解。

$$\hat{\beta} = \left[\left(\sum_{i=1}^{N}x_i'z_i\right)\left(\sum_{i=1}^{N}z_i'z_i\right)^{-1}\left(\sum_{i=1}^{N}z_i'x_i\right)\right]^{-1}\left(\sum_{i=1}^{N}x_i'z_i\right)\left(\sum_{i=1}^{N}z_i'z_i\right)^{-1}\left(\sum_{i=1}^{N}z_i'y_i\right)$$

得到样本矩

$$\hat{\beta}^{IV} = \left(N^{-1}\sum_{i=1}^{N}z_i'x_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}z_i'y_i\right) = (Z'X)^{-1}Z'Y$$

代入 y_i，讨论渐进分布

$$\begin{aligned}\hat{\beta} = \beta + &\left[\left(N^{-1}\sum_{i=1}^{N}x_i'z_i\right)\left(N^{-1}\sum_{i=1}^{N}z_i'z_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}z_i'x_i\right)\right]^{-1}\\ &\cdot\left(N^{-1}\sum_{i=1}^{N}x_i'z_i\right)\left(N^{-1}\sum_{i=1}^{N}z_i'z_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}z_i'u_i\right)\end{aligned}$$

对应教材 5.2.2 Asymptotic Normality of 2SLS 讨论渐进分布和方差估计。

练习 5.7 秩条件的理解（额外） 证明 $rank(z'x)=k$ 和 $rank[E(x'z)E(z'z)^{-1}E(z'x)]=k$ 的充要性。

证明 [秩条件充要性]

第一部分：如果 $rank(z'x)=k$ 则 $rank[E(x'z)E(z'z)^{-1}E(z'x)]=k$。

因为 $rank(z'x)=k$ 讨论 $rank(z'x)v$

如果 $v\neq 0$，则 $rank(v'x)v\neq 0$

其中 $\hat{c}=X'(Z'Z)^{-1}X/N$ 以概率收敛于 c，且是对称满秩半正定矩阵，则一定是正定的（对角线元素不存在 0）。

$rank(c)=k$ 正定，则 $rank(\hat{c})$ 正定。

$E(z'z)$ 满秩正定

如果 $v\neq 0$，则 $v'E(z'x)E(z'z)^{-1}(xz')v>0$ 是对称矩阵，因此秩就是对角线个数。

第二部分：如果 $rank[E(x'z)E(z'z)^{-1}E(z'x)]=k$ 则 $rank(z'x)=k$。

如果 $v\neq 0$

$v'E(z'x)E(z'z)^{-1}E(xz')v\neq 0$

因为 $E(z'z)$ 是对称矩阵
如果 $v=0$
则前面的式子也为 0。
说明 $v\neq 0$，则 $E(z'x)v\neq 0$
第二部分，方法 2
已知 $rank(v'E(z'x)E(z'z)^{-1}E(xz')v)=k$, 拆解为 ABA 形式。

$$rank(ABA')=k$$

$$rank(A)\leq min\{K,L\}\leq K$$

$$rank(A)\leq min\{K,L\}\leq K$$

$$k=rank(ABA')=k\leq min(ABA')\leq rank(A')$$

5.4 工具变量的选择

命题 5.6 (好的工具变量选择思路)

同伴效应、关系之关系、自然地理、遥远的历史变量。

命题 5.7 (一阶段回归)

在所有的估计中 (都是一致的), 采用所有外生变量的线性组合来作为内生变量的工具变量时, 估计结果是最有效的 (方差最小的)。(证明参见:Wooldridge ,2002a,p96-97)

命题 5.8 (iv 和 2sls 等价)

证明两阶段回归等价 iv:
估计

$$y_1=\beta_0+\beta_1 y_2+\beta_2 z_1+u$$

工具变量

$$y_2^*=\pi_0+\pi_1 z_1+\pi_2 z_2+\pi_3 z_3$$

也就是说一阶段

$$y_2=\pi_0+\pi_1 z_1+\pi_2 z_2+\pi_3 z_3+v$$

这时候，以下回归方程等价:
1、回归 y_2、$\hat{y_2}$、z_1
2、回归 y_2、y_2、$\hat{v}$（一阶段回归的残差）、z_1
从矩条件和 OLS 估计的一阶条件等价出发来证明。

证明 [方程的关系]

第一种情况的矩条件为：

$$\sum(y_1-\beta_0-\beta_1\hat{y}_2-\beta_2 z_1)=0$$
$$\sum\hat{y}_2(y_1-\beta_0-\beta_1\hat{y}_2-\beta_2 z_1)=0$$
$$\sum z_1(y_1-\beta_0-\beta_1\hat{y}_2-\beta_2 z_1)=0$$

第二种情况的矩条件为：

$$\sum(y_1-\alpha_0-\alpha_1 y_2-\alpha_2\hat{v}-\alpha_3 z_1)=0$$
$$\sum\hat{v}(y_1-\alpha_0-\alpha_1 y_2-\alpha_2\hat{v}-\alpha_3 z_1)=0$$
$$\sum y_2(y_1-\alpha_0-\alpha_1 y_2-\alpha_2\hat{v}-\alpha_3 z_1)=0$$
$$\sum z_1(y_1-\alpha_0-\alpha_1 y_2-\alpha_2\hat{v}-\alpha_3 z_1)=0$$

证明核心思路就是这两个矩条件能互相转化。

此时有残差关系

$\hat{v}=y_2-\hat{y}_2\Longrightarrow y_2=\hat{y}_2+\hat{v}$ 和 $E(\hat{v})=0$

代入第二种情况：

$$\begin{aligned}&y_1-\alpha_0-\alpha_1 y_2-\alpha_2\hat{v}-\alpha_3 z_1\\&=y_1-\alpha_0-(\alpha_1+\alpha_2)\hat{v}-\alpha_1\hat{y}_2-\alpha_3 z_1\\&=y_1-\alpha_0-\alpha_1\hat{y}_2-\alpha_3 z_1\end{aligned}$$

由于

$$\sum\hat{v}\hat{y}_2=0,\quad\sum\hat{v}z_1=0,\quad\sum\hat{v}=0$$

此时简化矩条件：

$$\begin{aligned}&y_2(y_1-\alpha_0-\alpha_1\hat{y}_2-(\alpha_1+\alpha_2)\hat{v}-\alpha_3 z_1)\\&=(\hat{y}_2+\hat{v})(y_1-\alpha_0-\alpha_1\hat{y}_2-(\alpha_1+\alpha_2)\hat{v}-\alpha_3 z_1)\\&=\hat{y}_2(y_1-\alpha_0-\alpha_1\hat{y}_2-(\alpha_1+\alpha_2)\hat{v}-\alpha_3 z_1)\\&=\hat{y}_2(y_1-\alpha_0-\alpha_1\hat{y}_2-\alpha_3 z_1)\end{aligned}$$

$$z_1(y_1-\alpha_0-\alpha_1\hat{y}_2-(\alpha_1+\alpha_2)\hat{v}-\alpha_3 z_1)=z_1(y_1-\alpha_0-\alpha_1\hat{y}_2-\alpha_3 z_1)$$

注[多个 iv] 多个 iv 和单个 iv 的证明一样，不同的只是 $\hat{y}_2$ 内容不同。

上述检验其实就是对照了 iv 估计的两种形式：

$$\beta_{iv}=\frac{Cov(z,y)}{Cov(z,x)}=\frac{Cov(\hat{x},y)}{Var((\hat{x}))}=\frac{\frac{Cov(z,y)}{Varz}}{\frac{Cov(z,x)}{Var(z)}}$$

5.5 假设检验

5.5.1 内生性检验

5.5.1.1 Hausman 检验

推荐参考IV-控制函数法-内生性和 Hausman 检验的原理解读

假设：

$$\begin{cases} H_0: \ \beta_{2sls} & = \beta_{ols} \\ H_1: \ \beta_{2sls} & \neq \beta_{ols} \end{cases}$$

构造统计量

$$\begin{aligned} t &= (\hat{\beta}_{2sls} - \hat{\beta}_{ols})/se(\hat{\beta}_{2sls} - \hat{\beta}_{ols}) \\ &= (\hat{\beta}_{2sls} - \hat{\beta}_{ols})/\left\{\left[se(\hat{\beta}_{2sls})\right]^2 - \left[se(\hat{\beta}_{ols})\right]^2\right\}^{\frac{1}{2}} \end{aligned}$$

分母部分的变形利用了同方差假设 $Var(\hat{\beta}_{2sls} - \hat{\beta}_{ols}) = Var(\hat{\beta}_{2sls}) - Var(\hat{\beta}_{ols})$

异方差情况：

$$H = (\hat{\beta}_{2sls} - \hat{\beta}_{ols})'[Var(\hat{\beta}_{2sls}) - Var(\hat{\beta}_{ols})]^{-1}(\hat{\beta}_{2sls} - \hat{\beta}_{ols})$$

如果是**一组参数**则构建卡方统计量。

5.5.1.2 通用检验

对于估计式：

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u$$

z_1 为内生变量，y_2 为外生变量。工具变量为 z_2, z_3。

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v$$

此时

$$Cov(y_2, u) = Cov(v, u) \neq 0$$

也就是说借助工具变量，可以把带来内生性问题的部分鬼纳入 v 之中，那么此时可以理解成以下形式：

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \rho v + e$$

解释变量的内生性问题转化为 v 是否存在影响。

5.5.2 弱工具变量

弱工具变量检验: 检验第一阶段回归中工具变量是否与内生解释变量显著相关。

因此, 可以构造 F 统计量 (多个工具变量) 或 t 统计量 (单个工具变量) 来进行弱工具变量检验。
目前有两个成熟的检验方法可以用于检验弱工具变量:
1. Hall et al. (1996)——Anderson canonical correlations likelihood-ratio test;
检验 iv 和 x 是否完全不相关。$H_0 : Cov(zx) = 0$
2. Stock & Yogo (2002)——Cragg-Donald chi-squared test$H_0 : Cov(zx) \longrightarrow 0$
3. overidentify，多个工具变量 $H_0 : Cov(z, u) = 0$
检验相关性的强弱。

5.5.3 过度识别约束

同方差假设下：
过度识别约束检验：

$$\begin{cases} H_0\text{: 所有工具变量外生/存在过度识别} \\ H_1\text{: 原假设不成立} \end{cases}$$

1. 在 2sls 回归中，获得最终残差 $\hat{u}_1$。
2. 再将残差对所有外生变量回归，获得 R_1^2。
3. 假设所有 iv 和 u_1 不相关，此时得到 $nR_1^2 \sim \chi_q^2$。
q 为 iv 数量减去内生解释变量数量。

练习 5.8 只有一个 iv 的情况下能检验吗?
不能检验。
此时第二步就是对 0 做回归，$R^2 = 0$。
但是多个 iv 的情况则不同，
例如有两个工具变量 z_2, z_3, 此时有

$$u_1 = \gamma_0 + \gamma_1 z_2 + \gamma_2 z_3 + v$$

此时虽然有矩条件 $Cov(z_2, u) = 0, Cov(z_3, u) = 0$
但是 $\gamma_1 = 0, \gamma_2 = 0$ 并不一定。因为此时是偏相关。
在恰好识别条件下 $\gamma_1 = \frac{Cov(u_1,x)}{Var(x)} = 0$

笔记 [为什么 $nR_1^2 \sim \chi_q^2$]

$$\hat{u}_i = Z_i'\gamma + e_i$$

$$R^2 = \frac{\hat{u}'_{fit}\hat{u}_{fit}}{\hat{u}'\hat{u}} = \frac{\hat{u}'P_Z\hat{u}}{\hat{u}'\hat{u}}$$

$\hat{u}'P_Z\hat{u}$ 服从卡方分布是因为中间的 p_z 为对称幂矩阵。

问题 5.1 分析下图5.1：

解

其中弱工具变量检验是 Weak identification F 统计量相关部分 (中间)

F 统计量 320 > 10% 的相关部分 IV size = 19，拒绝原假设 (弱工具变量)。则此时不存在弱工具变量问题。

过度识别的检验在底部。

```
                                                      Number of obs =     10146
                                                      F(  8,   3214) =    167.75
                                                      Prob > F      =    0.0000
Total (centered) SS     =  1273.159926                Centered R2   =    0.0517
Total (uncentered) SS   =  150356.7537                Uncentered R2 =    0.9920
Residual SS             =  1207.275071                Root MSE      =     .3449

------------------------------------------------------------------------------
    lnweek_hour |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
lnfama_transfer |  -.0228683    .005861    -3.90   0.000    -.0343556   -.0113809
         lnwage |   .0154261   .0035082     4.40   0.000     .0085501    .0223021
           male |   .0545668   .0071341     7.65   0.000     .0405842    .0685494
            age |  -.0043466   .0004751    -9.15   0.000    -.0052778   -.0034154
            edu |  -.0240378   .0012536   -19.17   0.000    -.0264948   -.0215808
          _cons |   4.160451   .0388584   107.07   0.000      4.08429    4.236612
------------------------------------------------------------------------------
Underidentification test (Anderson canon. corr. LM statistic):          320.564
                                                   Chi-sq(2) P-val =    0.0000
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):               165.397
Stock-Yogo weak ID test critical values: 10% maximal IV size              19.93
                                         15% maximal IV size              11.59
                                         20% maximal IV size               8.75
                                         25% maximal IV size               7.25
Source: Stock-Yogo (2005).  Reproduced by permission.
------------------------------------------------------------------------------
Sargan statistic (overidentification test of all instruments):            2.156
                                                   Chi-sq(1) P-val =    0.1421
------------------------------------------------------------------------------
```

图 5.1: 2SLS 回归结果

当工具变量的数量大于内生解释变量的数量时，我们称这个模型为过度识别模型。过度识别约束检验的主要目的是检验所有的工具变量是否都是有效的。如果所有的工具变量都是有效的，那么不同的工具变量估计出来的参数应该是相互印证的。如果有无效的工具变量，那么使用这个无效的工具变量估计出来的参数可能会与其他的估计结果有显著差别.

H0: 所有的工具变量都是有效的; **H1:** 至少有一个工具变量是无效的。

此时 P 值为 0.1421>0.05。因此接受原假设——所有工具变量都有效——不存在过度识别假设。

5.6 IV 与 ols 的系数比较

iv 往往会放大 ols 的系数。实际上是无偏性（也可以看作内部有效性）和外部有效性（结论外推）的取舍。

国际著名金融学家、RFS 主编、哥伦比亚大学商学院讲席教授姜纬 (Wei Jiang) 教授在《Have Instrumental Variables Brought Us Closer to the Truth?》一文中对 JF、JFE 和 RFS 三大金融顶级期刊中 255 篇运用工具变量进行回归估计的论文进行研究后发现，工具变量估计法将回归估计系数平均扩大了九倍。

参见：

1. Imbens G W, Angrist J D. Identification and Estimation of Local Average Treatment Effects[J]. Econometrica, 1994, 62(2):467-475.
2. Lecture from Imbens & Wooldrige: instrumental variables with treatment effect heterogeneity:local average treatment effect.(这个是在课程讲义中)

基于 LATE 视角，简单来说，可以把 iv 看作一种权重。强化了理想样本的权重（顺从者）。也就是强化了从 x 到 iv 中单调变化的数据权重。因此无偏，但外部有效性（推广到其他群体，例如面对政策试点反着来的人）降低了。

对于有内生性的问题，无论正负，我们都希望系数的绝对值被**低估**了。

因为说明即便低估了，模型依旧有显著性，真实情况会更好。

参考伍德里奇教材导论第六版 410 页。实际上可以如此表示 iv 和 ols 的渐近估计：

$$plim\tilde{\beta}_{IV} = \beta_1 + \frac{Corr(z,u)}{Corr(z,x)} \cdot \frac{\sigma_u}{\sigma_x}$$

$$plim\hat{\beta}_{OLS} = \beta_1 + Corr(x,u) \cdot \frac{\sigma_u}{\sigma_x}.$$

在美国，先抽签参军，此时每个人存在四种策略：

1. never-taker: 无论抽到参军还是不参军，都坚决不参军。
2. difier: 抽到参军就不去参军，抽到不参军就去参军。
3. complier：抽到参军就去参军，抽到不参军就不去参军。
4. always-taker: 无论抽到参军还是不参军，都坚决参军。

下面这个表格中右上角，策略表示 $W_i(0) = 1; W_i(1) = 0$。分别代表如果没抽中但会参军；如果抽中了就不去参军。

表 5.1: 抽签参军的四种策略

		$W_i(0)$	
		0	1
$W_i(1)$	0	*never-taker*	*defier*
	1	*complier*	*always-taker*

为了便于分析，此时假设没有人采用“抗逆者”（difier）策略，因为**抽签抽到参军资格,然后去参军的好处 ≥ 抽签没有抽到参军资格,然后去参军的好处**。

此时分布情况为：

表 5.2: 观察到的抽签和参军的群体

		Z_i	
		0	1
W_i	0	*complier/never-taker*	*never-taker/defier*
	1	*always-taker/defier*	*complier/always-taker*

此时分布情况为：

表 5.3: 观察到的抽签和参军的群体：没有抗逆者

		Z_i	
		0	1
W_i	0	*complier/never-taker*	*never-taker*
	1	*always-taker*	*complier/always-taker*

估计三种人群 complier、never-taker、always-taker 的概率分布，此时有

$$\pi_a + \pi_c + \pi_n = 1$$

由于抽签是随机的，因此可以得到 $Z_i \perp \{W_i(0), W_i(1)\}$。

对于 always-taker 来说

二元变量的期望就是其等于 1 的概率（二项分布）：

$$\pi_\alpha = Pr[W_i(0) = W_i(1) = 1] = E[W_i | Z_i = 0]$$

因为随机分配，只要通过随机抽样的方式抽中一个局部，满足条件的概率就是那个群体。此时表格5.3左下角部分就只有一个策略群体（always-taker），其满足条件的期望就是群体分布概率。在 01 分布的情况下，可以转化为期望。

对于 complier 来说：

此时锁定右下角的局部进行估计，里面同时包含了 complier/always-taker，再额外减去 always-taker 的概率。

$$\pi_c = Pr[W_i(0) = 0, W_i(1) = 1] = E[W_i|Z_i = 1] - E[W_i|Z_i = 0]$$

对于 never-taker 来说：

$$\pi_n = Pr[W_i(0) = W_i(1) = 0] = 1 - E[W_i|Z_i = 1]$$

推导出各个群体的概率分布后，继续表示 y 的条件概率。其实 y 括号内代表最终是否去参军的都状态。

$$E[Y_i|W_i = 0, Z_i = 0] = \frac{\pi_c}{\pi_c + \pi_n} E[Y_i(0)|complier] + \frac{\pi_n}{\pi_c + \pi_n} E[Y_i(0)|never - taker]$$

$$E[Y_i|W_i = 0, Z_i = 1] = E[Y_i(0)|never - taker]$$

$$E[Y_i|W_i = 1, Z_i = 0] = E[Y_i(1)|always - taker]$$

$$E[Y_i|W_i = 1, Z_i = 1] = \frac{\pi_c}{\pi_c + \pi_a} E[Y_i(1)|complier] + \frac{\pi_a}{\pi_c + \pi_a} E[Y_i(1)|always - taker]$$

最终可以得到：

$$\beta^{\mathrm{IV}} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[W_i|Z_i = 1] - E[W_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|complier]$$

$$E[Y_i|Z_i = 1] = (\pi_c + \pi_a) E[Y_i|W_i = 1, Z_i = 1] + \pi_n E[Y_i|W_i = 0, Z_i = 1]$$

$$E[Y_i|Z_i = 0] = \pi_a E[Y_i|W_i = 1, Z_i = 0] + (\pi_c + \pi_n) E[Y_i|W_i = 0, Z_i = 0]$$

✍ **练习 5.9 计量导论教材习题** 考虑简单回归模型

$y = \beta_0 + \beta_1 x + u$

令 z 为 x 的二值工具变量。运用 $\hat{\beta}_1 = \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})}$, 证明 IV 估计量 $\hat{\beta}_1$ 可以写成：

$$\tilde{\beta}_1 = (\bar{y}_1 - \bar{y}_0)/(\bar{x}_1 - \bar{x}_0)$$

其中，$\bar{y}_0$ 和 $\bar{x}_0$ 是 $z_i = 0$ 的那部分样本中 y_i 和 x_i 的样本均值，而可和 $\bar{x}_1$ 是 $z_i = 1$ 的那部分样本中 y_i 和 x_i 的样本均值。该估计量称为群组估计量，它由沃德 (Wald,1940) 最先提出。

证明 [证明如下（法 1）]

$$\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y}) = \sum_{z_i=0} (0 - \bar{z})(y_i - \bar{y}) + \sum_{z_i=1} (1 - \bar{z})(y_i - \bar{y})$$

消除 $z_i = 0$ 的部分：

$$\sum_{z_i=0}(-\bar{z})\left(y_i - \bar{y}\right) = -\bar{z}\sum_{z_i=0}(y_i - \bar{y}) = -\bar{z}\left(\sum_{z_i=0} y_i - n_0\bar{y}\right)$$

代入 $\sum_{z_i=0} y_i = n_0\bar{y}_0$，

$$-\bar{z}\left(n_0\bar{y}_0 - n_0\bar{y}\right) = n_0\bar{z}(\bar{y} - \bar{y}_0)$$

$z_i = 1$ 的部分：

$$\sum_{z_i=1}(1-\bar{z})\left(y_i - \bar{y}\right) = (1-\bar{z})\sum_{z_i=1}(y_i - \bar{y}) = (1-\bar{z})\left(\sum_{z_i=1} y_i - n_1\bar{y}\right)$$

代入 $\sum_{z_i=1} y_i = n_0\bar{y}_0$，

$$(1-\bar{z})\left(n_1\bar{y}_1 - n_1\bar{y}\right) = n_1(1-\bar{z})(\bar{y}_1 - \bar{y})$$

由于 z 为 01 变量，此时 $\bar{z} = \frac{n_1}{n}, 1 - \bar{z} = \frac{n_0}{n}$
分子化简为

$$\frac{n_0 n_1}{n}(\bar{y}_1 - \bar{y}_0)$$

同理，分母为

$$\frac{n_0 n_1}{n}(\bar{x}_1 - \bar{x}_0)$$

证明 [证明如下（法 2）]
由题意得：

$$\bar{y}_1 = E[y \mid z = 1], \quad \bar{y}_0 = E[y \mid z = 0], \bar{x}_1 = E[x \mid z = 1], \quad \bar{x}_0 = E[x \mid z = 0].$$

此时有：

$$E[y] = p\bar{y}_1 + (1-p)\bar{y}_0, \quad E[x] = p\bar{x}_1 + (1-p)\bar{x}_0$$

$$Cov(z, y) = E[zy] - E[z]E[y]$$

其中，由于 z 为 01 变量，因此

$$E[zy] = E[z \cdot E(y \mid z)] = p\bar{y}_1 E[z] = p$$

此时得到

$$Cov(z, y) = p\bar{y}_1 - p\left(p\bar{y}_1 + (1-p)\bar{y}_0\right) = p\left(\bar{y}_1 - p\bar{y}_1 - (1-p)\bar{y}_0\right) = p(1-p)\left(\bar{y}_1 - \bar{y}_0\right)$$

同理，

$$Cov(z, x) = p(1-p)\left(\bar{x}_1 - \bar{x}_0\right)$$

此时

$$\frac{\mathrm{Cov}(z, y)}{\mathrm{Cov}(z, x)} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0} = \beta_{iv}$$

第六章 面板固定随机效应

6.1 面板数据

之前的都是截面数据：同一时间，不同个体。面板数据多了**时间信息**。因此可以缓解很多内生性担忧。

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

个体效应 c_i 不随时间变动而变动。

大部分时候我们都使用大 N 小 T（短面板）而不是长面板（大 T 小 N），因为短面板更容易满足独立同分布使用大数定理和中心极限。而长面板往往为时间序列，存在序列相关。

6.2 假设

$$corr(c_i, u_{it}) = 0$$

$$E(x'_{it}u_{it}) = 0$$

c_i: 不可观测到的异质性。个体固定效应（不随时间变化而变化，所以脚标没有 t）

u_{it} 为特异性误差。

试图缓解 c_i 的内生性担忧。

6.3 随机效应

对于面板

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

其他假定和 OLS 相同，这里强调核心的假定：

严格外生性假设：

$$(\mathbf{a})E(u_{it}|x_i, c_i) = 0(t = 1, \cdots, T)$$

任何时期的 x 都要满足这个条件。$E(x'_{it}v_{is}) = 0, \forall s, t, \quad v_{is} = c_i + vis$ 所以是严格外生性。

$$(\mathrm{b})E(c_i|x_i) = E(c_i) = 0$$

使用 OLS 估计可知 $E(x'_{it}c_i) = 0和E(x'_{it}u_{it}) = 0$ 同期满足即可。

使用初级计量的估计，只是多了时间和个体的加总。$\sum_i \sum_t x_{it}$

OLS 估计如下

$$\hat{\beta} = \sum_i \sum_t (x_{it}x'_{it})x'_{it}y_{it}$$

可以看出在一致性的条条件下，实际上只需要同期外生。

$$plim\hat{\beta} = \beta + \sum_i \sum_t E(x'_{it}x_{it})E(x'_{it}v_{it})$$

其中 $v_{it} = c_i + u_{it}$

只满足同期外生而不是严格外生，此时估计是一致的，但是不是最有效的，因为方差包含了过去和未来的协方差：

考虑随机变量

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

对于 $v_{it} = c_i + u_{it}$，有 $E(x'_{it}v_{it}) = 0$。

因此在满足严格外生性的情况下：

$$Var(v_{it}) = \sigma_c^2 + \sigma_u^2$$
$$E(v_{is}v_{it}) = E[(c_i + u_{is})(c_i + u_{it})] = \sigma_c^2$$

其中，这里的 s 是指其他时期。

于是得到：

$$\Omega = E(v_i v_i') = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 & \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & & \sigma_c^2 & \\ & \vdots & & \ddots & \vdots \\ \sigma_c^2 & & \sigma_c^2 & \cdots & \sigma_c^2 + \sigma_u^2 \end{bmatrix} = \sigma_u^2 \cdot I_T + \sigma_c^2 \cdot j_T j_T'$$

$v_i = (v_{i1}, v_{i2}, \cdots, v_{iT})'$ 是一个 T×1 向量，j_T 是元素为 1 的列向量。

于是模型就可以被改为和截面估计类似的情况：

$$y_i = X_i\boldsymbol{\beta} + v_i$$

注[类似 GLS 章节3.8.1处理异方差] 接下来为了得到有效的估计，利用 Ω 进行变化，使得方差消除自相关，变成单位矩阵：因此就是 $Var(\Omega^{-\frac{1}{2}}v) = \Omega^{-1}Var(v)$

- **随机效应假定 1**：(a)$E(u_{it}|x_i, c_i) = 0 (t = 1, \cdots, T); (b)E(c_i|x_i) = E(c_i) = 0$
- **随机效应假定 2**：$Rank\ E(X_i'\widehat{\Omega}^{-1}X_i) = k$ （不存在完全多重共线性）
- **随机效应假定 3**：(a)$E(u_i u_i'|x_i, c_i) = \sigma_u^2 I_T$; (b)$E(c_i^2|x_i) = \sigma_c^2$

假定（1）的 a 意味着严格外生：$E(x'_{it}v_{is}) = 0, \forall s, t, v_{is} = c_i + u_{is}$

假定（1）的 b 意味着不存在个体效应导致的内生性问题。

一致性只需要同期外生假设 $E(x'_{it}c_i) = 0$和$E(x'_{it}u_{it}) = 0$，但此时不是有效的。

这里的假定（3）类似同方差假定，但是并非必须，可以放宽。

即可得到 $\Omega = \sigma_u^2 \cdot I_T + \sigma_c^2 \cdot j_T j_T'$

$$\widehat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i=1}^N X_i'\widehat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^N X_i'\widehat{\Omega}^{-1}y_i\right)$$

注[式子的理解] 注意

这里面的向量 X_i 包含了对 t 的累加。

类似工具变量加权，对估式整体乘以 $\Omega^{-1/2}$ 进行加权，最终结果就是上面的估计式。

$$\Omega^{-1/2}y_i = \Omega^{-1/2}X_i\boldsymbol{\beta} + \Omega^{-1/2}v_i$$

注[一致性的条件] 满足一致性的关键在于假定 1。假定 1 不满足时，这时候 y_i 部分会出现所有时期的 u_{it}，如果不满足严格外生性，不一定为一致的。Ω 的所有时期都会和 u_{it} 相乘。

接下来关键就是估计 $\hat{\Omega}$

由于

$$Var(v_{it}) = \sigma_c^2 + \sigma_u^2$$
$$E(v_{is}v_{it}) = E[(c_i + u_{is})(c_i + u_{it})] = \sigma_c^2$$

注[理解方差] σ_c^2 就是 v_{is} 和 v_{it} 的协方差。其中，这里的 s 是指其他时期。

1. OLS 估计等式$y_i = X_i\boldsymbol{\beta} + v_i$，获得参数估计值$\widetilde{\boldsymbol{\beta}}$及残差$\widetilde{\boldsymbol{v}}_{it}$；
2. 计算$\hat{\boldsymbol{\delta}}_v^2 = (NT-k)^{-1}\sum_i\sum_t \hat{\boldsymbol{v}}_{it}^2$；
3. 计算$\hat{\boldsymbol{\sigma}}_c^2 = (NT(T-1)/2-k)^{-1}\sum_t\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\widetilde{\boldsymbol{v}}_{is}\widetilde{\boldsymbol{v}}_{it}$;(类似于简单线性回归模型中$\hat{\boldsymbol{\sigma}}^2$的估计)
4. 计算$\hat{\boldsymbol{\omega}}^2 = \hat{\boldsymbol{\sigma}}_v^2 - \hat{\boldsymbol{\sigma}}_c^2$；
5. 构造$\widehat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\sigma}}_u^2 \cdot l_T + \hat{\boldsymbol{\sigma}}_c^2 j_T j_T'$

异方差时，可以直接计算：$\tilde{\Omega} = N^{-1}\sum_{i=1}^{N}\tilde{v}_i\tilde{v}_i'$

6.4 固定效应

固定效应假定 1：(a)$E(u_{it}|x_i, c_i) = 0(t = 1, \cdots, T)$ 但是允许 $E(c_i|\boldsymbol{x_{it}}) \neq 0$

由于 $E(c_i|\boldsymbol{x_{it}}) \neq 0$，需要通过变换去除，但是这种处理建立在线性模型的假定基础之上。

- 组内均值变换 (Fixed Effect,FE)
- 差分变换 (First Difference,FD)
- 向前正交变换 (Forward Orthogonal Deviations, Arellano and Bover, 1995)(了解)

注[固定效应的理解] 固定效应在放松假设后，x_{it} 为内生变量，遗漏变量 c_i。思想就是通过变化除去 c_i 的影响。因此固定效应如何变换是不唯一的，有多种方法的，我们常用的一般都是组内均值变换。

变换可能带来系数含义的改变。

组内变换：Between **Equation**：时间维度上取得组内均值。

$$\bar{y}_i = \bar{x}_i\beta + c_i + \bar{u}_i$$

在这个式子中：$\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}, \bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it}$ 且$\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$

Within Equation：

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + u_{it} - \bar{u}_i$$

变形可得：

$$\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it}$$

固定效应假定 2：$Rank(\sum_{t=1}^{T} E(\ddot{x}_{it}'\ddot{x}_{it})) = k$

固定效应假定 3：$E(u_i u_i'|x_i, c_i) = \sigma_u^2 I_T$

$$\widehat{\beta}_{FE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}'\ddot{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}'\ddot{y}_{it}\right)$$

✍ **练习 6.1 理解假设** 固定效应满足一致性需要什么条件

解 同理，此处一致性依旧需要严格外生，因为 $\ddot{u}_{it}$ 也包含了所有时间个体的残差。

这里强调这一点是因为固定效应三个假定中，假定 2 是对 $\ddot{x}$ 进行假定，而假定 1 直接对 x_{it} 进行假定。和随机变量的一致性相同，此时需要严格外生性。

6.4.1 FE 组内变换求解固定效应

接下来考虑渐进分布：

对于

$$y_i = X_i\beta + c_i j_T + u_i$$

同样定义投影矩阵和消失矩阵

$$Q_T = I_T - j_T(j_T' j_T)^{-1} j_T' = I_T - \frac{1}{N} j_T j_T'$$

$$Q_T j_T = 0, Q_T h_i = \ddot{h_{it}}$$

此时投影矩阵就是组内去均值的操作含义，因此，估计可以看作两边同时乘以投影矩阵：

$$Q_T y_i = Q_T X_i \beta + Q_T u_i$$

注 因此，虽然转化为了 $\ddot{x}$，但假定 3 同方差假定依旧是在原方程基础上假设的。

同时，如果对单位矩阵回归，单位矩阵代表变量个数，看作常数，回归得到的其实是均值。β=y/N。

$j_T(j_T'j_T)^{-1}j_T'y_i = j_T((j_T'j_T)^{-1}j_T'y) = \bar{y}$

虚拟变量基于 FWL 定理表示，本质和组间估计形式相同。本质就是对 T 时间内同一个体的均值做离差（$x_i - \bar{x}_i$）。

得到估计式

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i d_i + u_{it}$$

回归 x_{it} 和 d_i, 得到 $\ddot{x} = M_1 d_i$

回归 y_{it} 和 x_{it}, 得到 $\ddot{y} = M_1 x_{it}$

回归 $\ddot{y}$ 和 $\ddot{x}$，本质是式子两侧同时乘以 $M_i = Q_T$

渐进方差为：

$$\sqrt{N}(\hat{\beta}_{FE} - \beta) = \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{X}_i'\ddot{X}_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ddot{X}_i'\ddot{u}_i\right)$$
$$= \left(\frac{1}{N}\sum_{i=1}^{N}\ddot{X}_i'\ddot{X}_i\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ddot{X}_i' u_i\right)$$

同方差假定 $E\left(u_i u_i'|\ddot{X}_i\right) = \sigma_u^2 I_T$ 成立时，可以得到简化形式

$$Var(\hat{\beta}_{FE}) = \sigma_u^2 \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i \right)^{-1}$$

$$\begin{aligned} E\left(\ddot{u}_{it}^2\right) &= E[(u_{it} - \bar{u}_i)^2] = E\left(u_{it}^2\right) + E\left(\bar{u}_i^2\right) - 2E(u_{it}\bar{u}_i) \\ &= \sigma_u^2 + \frac{\sigma_u^2}{T} - 2\frac{\sigma_u^2}{T} = \sigma_u^2\left(1 - \frac{1}{T}\right) \end{aligned}$$

对 t 求和

$$\sum_{t=1}^{T} E\left(\ddot{u}_{it}^2\right) = \sigma_u^2 (T-1)$$

对 n 求和

$$[N(T-1)]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E(\ddot{u}_{it}^2) = \sigma_u^2 E\left([N(T-1)]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{u}_{it}^2\right) = \sigma_u^2$$

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)-k} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2$$

注[自由度的理解] 这里减去 k 不影响一致性，但能保证无偏性。每个组内估计消耗了一段信息。这里的 N(T-1)-K 中，N 个个体，每个消耗掉了一个时间期内的均值，因此是 N(T-1)，然后又消耗掉了 K 个解释变量，因此自由度为 N(T-1)-K。

异方差时，估计量为：

$$\widehat{Var(\hat{\beta}_{FE})} = \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \ddot{X}_i' \widehat{u}_i \widehat{u}_i' \ddot{X}_i \right) \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i \right)^{-1}$$

✍ **练习 6.2 思考** 思考 1：如果 $Cov(u_{it}, u_{is}) \neq 0 (t \neq s)$, 即 $E(u_i u_i' | x_i, c_i) = \Lambda$ 非对角线上的元素可能不为 0。那么，如何在固定效应估计的基础上使用 FGLS 估计？(P276-277)

解 此时，直接构造 $\widehat{\Omega} = N^{-1} \sum_{i=1}^{N} \widehat{u}_i \widehat{u}_i'$ 不是一个可逆矩阵，可将每个 i 的某一期去掉，然后再在组内去均值的方程左乘 $\widehat{\Omega}^{-1/2}$。

因为此时矩阵为 $Q_T u_i$ 构造出的矩阵，此时秩为 $k-1$，最终需要再去除每个 i 的某一期。

✍ **练习 6.3 思考** 思考 2: 因为 $E(\ddot{u}_{it}^2) = \sigma_u^2 \left(1 - \frac{1}{T}\right)$, 那么, 可否先基于方程 $\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it}$ 估计 $E(\ddot{u}_{it}^2)$, 再将 $E(\ddot{u}_{it}^2)$ 的估计量乘以 $\frac{T}{T-1}$ 作为 σ_u^2 的估计量?

意思就是我们现在已经估计出方差：

$$\hat{\sigma}_{\ddot{u}}^2 = \frac{1}{N(T-1)-k} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2 \left(\hat{u}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\widehat{\boldsymbol{\beta}}_{\boldsymbol{FE}}\right) \tag{6.1}$$

但是这个估计量为 $\ddot{y}_{it}$ 方程的估计量而不是方程 y_{it} 的估计量。此时是否需要再次乘以 $\frac{T}{T-1}$ 还原回原方程的估计量。

解 不会，且此时会高估标准误。因为自由度是在最后调整的，式子6.1中自由度 $-k$ 那一步就是调整的过程了。

✍ **练习 6.4 思考** 思考 3: 如何利用固定效应模型估计 σ_c^2? (注意: 固定效应模型已经消除了 c_i)

解 一旦获得 $\widehat{\boldsymbol{\beta}}_{FE}$, 我们便可计算原始模型的结构误差 $v_{it} = c_i + u_{it}$ 的估计量 $\widehat{v}_{it} = y_{it} - x_{it}\widehat{\boldsymbol{\beta}}_{FE}$, 由此便可得到 $\widehat{\sigma}_v^2$, 利用 $\widehat{\sigma}_v^2 = \widehat{\sigma}_c^2 + \widehat{\sigma}_u^2$ 便可计算出 $\widehat{\sigma}_c^2$。

✍ **练习 6.5 思考** 固定效应模型中估计的 $\widehat{\sigma}_c^2$ 通常大于随机效应。(为什么?)。

解 随机效用控制的东西更多，而固定效应包含了一切不随时间变化而变化的要素。

在假设中，随机效应假设 a_i 和 x_i 不相关。因此把部分个体效应差异解释为了 x，此时随机效应会低估方差。从数学上看，就是权重矩阵拉伸标准化了这部分异方差。

固定效应允许 a_i 和 x_i 相关，因此 a_i 相较而言解释了更多的个体差异。

也就是，固定效应在分解去均值时，去除的那部分个体固定效应比随机效应更大，更直观的表述可以参考练习6.6

6.4.2 LSDV 虚拟变量求解固定效应

基于个体生成虚拟变量参与回归。在固定效应假设满足时，结果和组间变换计算的个体固定效应等价。（证明:Green, 7th edition, pp.360-361）

此时自由度 N(T-1)-K 可以理解为 NT 个样本，减去 K 个解释变量，减去 N 个虚拟变量解释变量定义。

6.4.3 FD 差分求解固定效应

使用一阶差分估计固定效应

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})\boldsymbol{\beta} + u_{it} - u_{it-1}$$

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it}$$

- 假定 1$E\left(u_{it}|x_i, c_i\right) = 0$。
  这一假定可以保证一阶差分估计量是无偏的，实际上，保证估计一致性只需要 $\boldsymbol{E}(\Delta \boldsymbol{x}_{it}' \Delta u_{it}) = 0$。理解上，随机效应、fe、lsdv 的固定效应计算，一致性需要满足严格外生性，这里使用的是差分，相当于满足了严格外生。
- 假定 2$Rank(\sum_{t=2}^{T} E(\Delta x_{it}' \Delta x_{it})) = k$
- 假定 3$E(\Delta u_i' \Delta u_i | x_i, c_i) = \sigma_{\Delta u}^2 I_{T-1}$。$\Delta u_i$ 同方差且序列无关。

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}'\Delta x_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}'\Delta y_{it}\right)$$

同方差下，

$$Var(\hat{\beta}_{FD}) = \sigma_{\Delta u}^2\left(\sum_{i=1}^{N}\Delta X_i'\Delta X_i\right)^{-1}$$

$$\hat{\sigma}_{\Delta u}^2 = \frac{1}{N(T-1)-k}\sum_{i=1}^{N}\sum_{t=2}^{T}\widehat{\Delta u}_{it}^2$$

异方差下，

$$Var\widehat{\left(\hat{\beta}_{FD}\right)} = \left(\sum_{i=1}^{N}\Delta X_i'\Delta X_i\right)^{-1}\left(\sum_{i=1}^{N}\Delta X_i'\widehat{\Delta u}_i\widehat{\Delta u}_i'\Delta X_i\right)\left(\sum_{i=1}^{N}\Delta X_i'\Delta X_i\right)^{-1}$$

6.5 效应的选择

6.5.1 FE 和 FD 的关系

✍ **练习 6.6 关系** 为什么我们现在都默认使用固定效应。再次对比理解下：

解 [关系] 面板数据中，都分离出了个体固定效应 a_i 和 u_{it}。

此时误差项为 $v_{it} = a_i + u_{it}$.

随机效应假设 a_i 完全随机，因此和 x_i 不相关，只是每个个体的误差项 v_{it} 都共享一个 a_i，在同一个个体之间的误差存在相关性（组内序列相关）。

而固定效应允许了 a_i 与 x_i 相关，相当于是放松了假设。然后想办法减去 a_i 的影响。

事实上，对于 ols 来说：

$$y = \beta x + u$$

随机效应就是

$$y - \theta\bar{y} = \beta(x - \theta\bar{x}) + (u - \theta\bar{u})$$

$\theta = 0$ 时，退化到 ols；$\theta = 1$ 时，退化到固定效应。而随机效应时，根据数据情况对方差进行拉伸（标准化），使得其变得更加有效。

我们更希望在正确的估计下做出保守的估计，因此一般都选择固定效应。随机效应要求更严格一些。OLS、re 成立，则 fe 一定成立。

✍ **练习 6.7 练习** 证明两期面板，FE=FD

解 两期面板：FE=FD（考过）

$$\widehat{\beta}_{FE} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}'\ddot{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}'\ddot{y}_{it}\right)$$

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}'\Delta x_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}'\Delta y_{it}\right)$$

$$\sum_{t=1}^{2}\ddot{x}_{it}'\ddot{x}_{it} = \sum_{t=1}^{2}\ddot{x}_{it}^2 = \sum_{t=1}^{2}(x_{it} - \bar{x}_i)^2$$

在两期中，组间均值为 $\bar{x}_i = \frac{x_{i1}+x_{i2}}{2}$

$$\begin{aligned}\sum_{t=1}^{2}(x_{it} - \bar{x}_i)^2 &= \sum_{t=1}^{2}(x_{it} - \frac{x_{i1}+x_{i2}}{2})^2 \\ &= \frac{(x_{i2}-x_{i1})^2}{2}\end{aligned}$$

对于 FD，则只有一期

$$\sum_{t=2}^{2}\Delta x_{it}'\Delta x_{it} = (x_{i2} - x_{i1})^2$$

多期面板，

u_{it} 序列无关，则 FE 更有效。

u_{it} 随机游走，则 FD 更有效。

如何证明两期面板中方差也相等。

命题 6.1 (固定效应中 FE 和 FD 的关系)

如下

两期面板 (T=2): FE=FD

多期面板 (T>2):

假如 u_{it} 序列无关，FE 估计量比 FD 估计量更有效；

假如 u_{it} 服从随机游走，则 FD 估计量比 FE 估计量更有效。

注[理解] 如下

随机游走时，上一时期和下一时期相关 $u_{it} = u_{i,t-1} + e_{it}$， $e_{it} \sim iid$，差分能减轻这部分相关。

序列无关时，差分相当于包含了两期误差。

$$\Delta u_{it} = u_{it} - u_{i,t-1}$$

$$Var(\Delta u) = Var(u_t) + Var(u_{t-1}) = 2\sigma^2$$

选择 FD 还是 FE

Hausman 检验:Wooldridge, 2005, problem 10.6。

6.6 检验

6.6.1 方法 1

豪斯曼检验是 wald 检验下的一类思想，系统检验估计量的差异。
都是构造如下形式的估计量。

$$H=(\hat{\beta}_{FE}-\hat{\beta}_{RE})'\left(\mathbf{Var}(\hat{\beta}_{FE}-\hat{\beta}_{RE})\right)^{-1}(\hat{\beta}_{FE}-\hat{\beta}_{RE})$$

由于固定效应方差更大（原因见思考题6.5），一般放在前面。
此时，

$$\mathbf{Var}(\hat{\beta}_{\mathrm{FE}}-\hat{\beta}_{\mathrm{RE}})=\mathbf{Var}(\hat{\beta}_{\mathrm{FE}})+\mathbf{Var}(\hat{\beta}_{\mathrm{RE}})-2\cdot\mathbf{Cov}(\hat{\beta}_{\mathrm{FE}},\hat{\beta}_{\mathrm{RE}})$$

此时，Hausman 检验[1]的原假设为（满足随机效应假设）：

$$H_0:\mathbf{Cov}(x_{it},a_i)=\mathbf{0}$$

此时基于效率定理——不同检验估计损失的效率和效率无关。

$$\mathrm{Cov}(\hat{\beta}_{\mathrm{RE}},\hat{\beta}_{\mathrm{FE}}-\hat{\beta}_{\mathrm{RE}})=\mathbf{0}$$

其**正交性**可以看作来自以下分解（依照原假设，RE 估计更有效）

$$\hat{\beta}_{\mathrm{FE}}=\hat{\beta}_{\mathrm{RE}}+(\hat{\beta}_{\mathrm{FE}}-\hat{\beta}_{\mathrm{RE}})$$

展开也就得到了

$$\mathbf{Cov}(\hat{\beta}_{\mathrm{RE}},\hat{\beta}_{\mathrm{FE}})-\mathbf{Var}(\hat{\beta}_{\mathrm{RE}})=\mathbf{0}$$

效率定理也揭示了对于两个估计量构成的协方差，结果为更有效的一方的方差。
此时代入得到：

$$\begin{aligned}\mathbf{Var}(\hat{\beta}_{\mathrm{FE}}-\hat{\beta}_{\mathrm{RE}})&=\mathbf{Var}(\hat{\beta}_{\mathrm{FE}})+\mathbf{Var}(\hat{\beta}_{\mathrm{RE}})-2\cdot\mathbf{Cov}(\hat{\beta}_{\mathrm{FE}},\hat{\beta}_{\mathrm{RE}})\\&=\mathbf{Var}(\hat{\beta}_{\mathrm{FE}})+\mathbf{Var}(\hat{\beta}_{\mathrm{RE}})-2\cdot\mathbf{Var}(\hat{\beta}_{\mathrm{RE}})\\&=\mathbf{Var}(\hat{\beta}_{\mathrm{FE}})-\mathbf{Var}(\hat{\beta}_{\mathrm{RE}})\end{aligned}$$

最终，豪斯曼检验化简为：

$$H=(\hat{\beta}_{FE}-\hat{\beta}_{RE})'\left[Var\widehat{(\hat{\beta}_{FE})}-Var\widehat{(\hat{\beta}_{RE})}\right]^{-1}(\hat{\beta}_{FE}-\hat{\beta}_{RE})$$

式子服从 χ 分布。此时基于中心极限定理，有渐进正态分布 $\hat{\beta}_{FE}-\hat{\beta}_{RE}\sim N(\mathbf{0},\mathbf{\Sigma})$。标准化之和为卡方分布。

单个检验所转化为 t 检验。$\hat{\beta}_{FE}-\hat{\beta}_{RE}$ 标准化后作为分子，分母为标准差 $t=\frac{z-0}{\sigma}$。

$$t=(\hat{\beta}_{FE}-\hat{\beta}_{RE})/\left\{\left[se(\hat{\beta}_{FE})\right]^2-\left[se(\hat{\beta}_{RE})\right]^2\right\}^{\frac{1}{2}}$$

✍ **练习 6.8 基于同方差证明？** 采用同方差思路证明？

$$Var(\beta_{RE})=\sigma^2(\Omega^{\frac{1}{2}}X')^{-1}(\Omega^{\frac{1}{2}}X)^{-1}=\sigma^2A'A$$

[1] 文献原文Specification Tests in Econometrics

$$Var(\beta_{FE}) = \sigma^2 (MX')^{-1}(MX)^{-1} = \sigma^2 B'B$$

此时

$$Var(\beta_{FE} - \beta_{RE}) = \sigma^2 (B'B + A'A - A'B - B'A)$$

想要证明 $Var(\beta_{FE} - \beta_{RE}) = Var(\beta_{FE}) - Var(\beta_{RE})$ 转化为了证明

$$B'B + A'A - A'B - B'A = B'B - A'A$$

也就是

$$A'B + B'A = 2A'A$$

此时将系数分解为

$$\hat{\beta}_{FE} = \hat{\beta}_{RE} + (\hat{\beta}_{FE} - \hat{\beta}_{RE})$$

$$\mathrm{Var}(\hat{\beta}_{FE}) = \mathrm{Var}(\hat{\beta}_{RE}) + \mathrm{Var}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) + \mathrm{Cov}(\hat{\beta}_{RE}, \hat{\beta}_{FE} - \hat{\beta}_{RE}) + \mathrm{Cov}(\hat{\beta}_{FE} - \hat{\beta}_{RE}, \hat{\beta}_{RE})$$

若原假设成立，随机效应更优，则随机效应更有效，方差更小

$$\mathrm{Var}(\hat{\beta}_{FE}) = \mathrm{Var}(\hat{\beta}_{RE}) + \mathrm{Var}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) + \mathrm{Cov}(\hat{\beta}_{RE}, \hat{\beta}_{FE} - \hat{\beta}_{RE}) + \mathrm{Cov}(\hat{\beta}_{FE} - \hat{\beta}_{RE}, \hat{\beta}_{RE}) \geqslant \mathrm{Var}(\hat{\beta}_{RE})$$

$$\mathrm{Var}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) + \mathrm{Cov}(\hat{\beta}_{RE}, \hat{\beta}_{FE} - \hat{\beta}_{RE}) + \mathrm{Cov}(\hat{\beta}_{FE} - \hat{\beta}_{RE}, \hat{\beta}_{RE}) \geqslant 0$$

Listing 6.1: 检验

```
xtreg y x, fe est store fe

xtreg y x, re est store re

hausman fe re, sigmamore/sigmaless
```

6.6.2 方法 2

ols 和随机效应对一致性的要求并不相同。

$$\text{Pooled OLS:} \quad E(x_{it}' v_{it}) = 0, \text{同期外生。}$$
$$\text{Random Effect:} \quad E(x_{is}' v_{it}) = 0, \forall s, t, \text{严格外生。}$$

存在随机效应可以通过检验下列式子得到：

$$E(v_{is} v_{it}) = \sigma_c^2, \forall s \neq t$$

$$\hat{\sigma}_c^2 = (NT(T-1)/2-k)^{-1}\sum_i\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\tilde{v}_{is}\tilde{v}_{it}$$

对应的中心极限

$$\sqrt{N}\hat{\sigma}_c^2 \sim Normal\left(0, E\left[\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\tilde{v}_{is}\tilde{v}_{it}\right)^2\right]\right)$$

stata 命令

Stata 默认采用 Breusch-Pagan LM test 来检验是采用 pooled OLS 还是用 Random effects model, 拒绝原假设, 应采用 random effects model。

$$\text{随机效应存在} \rightarrow E(v_{is}v_{it}) = \sigma_c^2, \forall s \neq t$$

原假设对应方差不为 0。应用时会使用中心极限定理转化一下。

$$\sqrt{N}\hat{\sigma}_c^2 \sim Normal\left(0, E\left[\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\tilde{v}_{is}\tilde{v}_{it}\right)^2\right]\right)$$

Listing 6.2: 检验

```
xtreg y x, re
xttest0
```

Stata 默认采用 Breusch-Pagan LM test 来检验是采用 pooled OLS 还是用 Random effects model, 拒绝原假设, 应采用 random effects model。

6.6.3 方法 3

基于回归的检验。

若 $E(c_i|x_i) \neq 0$，则可以将 c_i 看成是 x_i 的函数。例如，可以令

$E(c_i|x_i) = E(c_i|w_i) = \gamma_0 + \overline{w}_i\gamma$

例如研究饮食习惯，无法直接估计个体效应，使用其每月水果中香蕉占比作为 c_i 的代理变量。

w_i 是 x_i 的子集，$\overline{w}_i$ 是 $\overline{x}_i$ 的子集，实践中可以用 $\overline{w}_i$ 作为 w_i 的替代。

在原假设成立的条件下，组内去均值的变量应对随机效应变换后的因变量无显著影响（原假设是 RE 方程为真），若有影响，遗漏 c_i（即 $\overline{w}_i$）将导致 β 估计有偏。因此，可以构造如下回归方程来检验:

$$\breve{y}_i = \breve{X}_i\beta + \overline{w}_i\xi + \breve{v}_i$$

- 其中，$\breve{h}_i = \widehat{\Omega}^{-\frac{1}{2}}h_i$，$\overline{w}_i$ 是 $\overline{x}_i$ 的子集。
- 检验 $\xi = 0$ 来判断是采用随机效应还是固定效应，拒绝原假设应采用固定效应。
- 也可以把 $\overline{w}_i$ 换为 $\breve{w}_i$，结果是等价的（Mundlak, 1978）。
- 此方法本质上是检验 c_i 是否影响 $\breve{y}_i$。隐含的假设为 c_i 是 $\overline{x}_i$ 的函数。

第七章 面板内生性

7.1 内生性来源

无论是 RE,FE,FD 还是 FOD, 都需要满足严格外生性假定一致性条件：$E(u_{it}|x_i, c_i)$
违背严格外生性就是违背此假定：$E(u_{it}|x_i, c_i) = 0(t = 1, \cdots, T)$

7.2 工具变量

此时, 要获得参数 β 的一致估计需要工具变量 Z,Z 满足外生性条件 (矩条件)。在面板数据中, 外生性矩条件可以分成四种类型:
$g_{L\times 1} = E(Z_i' v_i) = T^{-1}\sum_{t=1}^{T} z_{it}' v_{it} = 0($ 可加外生假定)
$E(z_{it} v_{it}) = 0($ 同期外生假定)
$E(z_{it} v_{is}) = 0, \forall t \leq s($ 序列外生假定)
$E(z_{it} v_{is}) = 0, \forall t, s($ 强外生假定 = 严格外生)

7.2.1 随机效应

此处 $Z_i'\Omega^{-1/2}$ 默认了工具变量数量和内生变量数量相等。

$$\beta_{REIV} = E(Z_i'\Omega^{-1}X_i)^{-1}E(Z_i'\Omega^{-1}y_i)$$

$$\hat{\beta}_{REIV} = \left(\sum_{i=1}^{N} Z_i'\Omega^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} Z_i'\Omega^{-1}y_i\right)$$

方差为：

$$\operatorname{var}(\hat{\beta}_{REIV}) = \left(\sum_{i=1}^{N} Z_i'\Omega^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} Z_i'\Omega^{-1}Z_i\right)\left(\sum_{i=1}^{N} X_i'\Omega^{-1}Z_i\right)^{-1}$$

7.2.2 固定效应

强外生假定满足 $E\left(\ddot{Z}_i'\ddot{u}_i\right) = 0(\ddot{Z}_i = Q_T Z_i)$
同理得到工具变量：

$$\operatorname{var}(\hat{\beta}_{FEIV}) = \left[\left(\sum_{i=1}^{N} \ddot{X}_i'\ddot{Z}_i\right)\left(\sum_{i=1}^{N} \ddot{Z}_i' u_i u_i' \ddot{Z}_i\right)^{-1}\left(\sum_{i=1}^{N} \ddot{Z}_i'\ddot{X}_i\right)\right]^{-1}$$

7.3 序列外生的动态面板

序列外生：y_t 与过去的 y_{t-1} 的关系相关。
对于动态面板：

$$y_{it} = x_{it}\beta + c_i + u_{it}$$

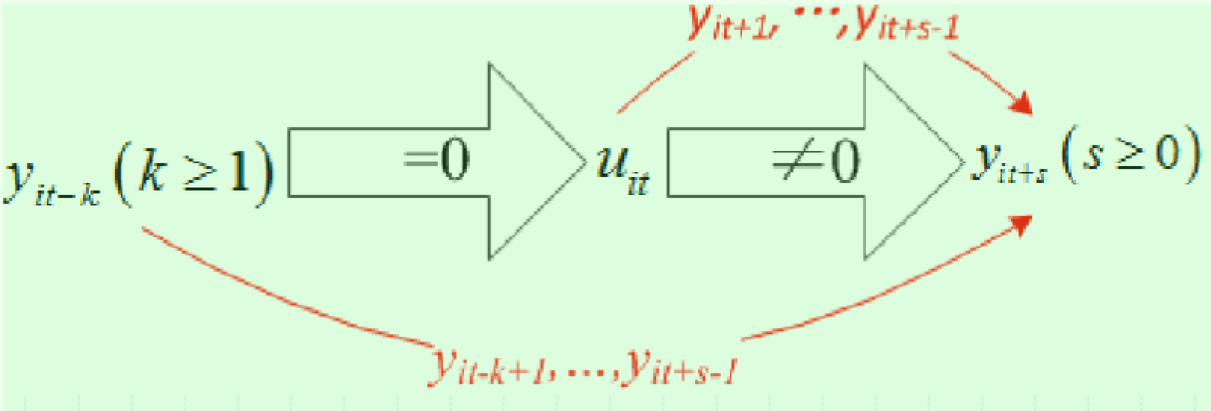


图 7.1: 动态完备

由于 x_{it} 和 u_{is} 很难满足严格外生性假设，因此做出如此假设：

$$E(u_{is} \mid x_{it}, c_i) = 0 \Rightarrow E(x'_{it}u_{is} \mid c_i) = 0, \forall t \leq s$$

定义 7.1

动态完备：过去的所有要素影响现在，排除未来的因素。同时既然每一个要素都是通过过去的要素影响，因此前一期即可作为核心解释变量。♣

$$E(y_{it} \mid x_{it}, x_{it-1}, ..., x_{i1}, c_i) = E(y_{it} \mid x_{it}, c_i) = x_{it}\beta + c_i$$

因此上述假设也称为序列 (序贯) 外生假设 (Sequentially Exogeneous)。
对应的有如下面板构造：

$$y_{it} = \rho y_{it-1} + c_i + u_{it}$$

假设满足 $x_{it} = y_{it-1}$, 且$E(y_{it-1}c_i) \neq 0$。
动态完备和之前相同：

$$E(y_{it}|y_{it-1}, ..., y_{i0}, c_i) = E(y_{it}|y_{it-1}, c_i) = \rho_1 y_{it-1} + c_i$$

u_{it} 序列不相关。$E(y_{it-k}u_{it}) = 0 \Rightarrow E(u_{it-k}u_{it}) = 0(k \geq 1)$
u_{it} 影响将来的 y_{it}。考虑到一阶自相关。$E(y_{it}u_{it-k}) = \rho^k E(y_{it-k}u_{it-k}) \neq 0, \forall k \geq 1$。
y_{it} 存在持续影响。$Cov(y_{it}, y_{is}) \neq 0$。
动态完备，则一定序列外生, 反之不成立。因为序列外生的变量不一定影响 y。
总结。动态完备中，u_{it}、y_{it-s} 只影响未来的 y_{it+1}；
当被解释变量存在动态反馈时,u_{it} 一定不满足严格外生假定, 如上例中,$E(y_{it}u_{it}) \neq 0$。
当被解释变量存在动态反馈时,c_i 一定和解释变量相关, 如上例中,$E(y_{it-1}c_i) \neq 0$。
动态完备模型中, 天然提供了外生的矩条件, 如上例中,$E(y_{it}u_{is}) = 0(\forall t < s)$

7.3.1 FD

FE 无法满足一致性，因为此时不满足外生性：

$$E[(y_{it-1}-\bar{y}_i)(u_{it}-\bar{u}_i)]\neq 0$$

以一阶差分为例子：

$$\Delta y_{it}=\rho\Delta y_{it-1}+\Delta u_{it}$$

所以都使用 FD 估计

$$E(\Delta y_{it-1}\Delta u_{it})=E[(y_{it-1}-y_{it-2})(u_{it}-u_{it-1})]=-\sigma_u^2\neq 0$$

OLS 与 FE 分别提供了参数 ρ 的上界和下界!(Bond,2002)
1. 借助 FD 方法去除 c_i 的影响;
2. 采用工具变量获得 ρ 的一致估计。
构建矩阵

$$D_{(T-2)\times(T-1)}=\begin{bmatrix}1 & -1 & 0 & \cdots & 0\\ 0 & 1 & -1 & & \\ \vdots & & \ddots & \ddots & \\ 0 & & & 1 & -1\end{bmatrix}$$

得到

$$Dy_i=\rho Dy_{i,-1}+DX_i\gamma+Du_i$$

$$y_i=(y_{i2},\cdots,y_{iT})',y_{i,-1}=(y_{i1},\cdots,y_{iT-1})'$$

$$Z_i=\begin{bmatrix}y_{i1}\\ \vdots\\ y_{T-1}\\ y_{iT-2}\end{bmatrix}_{(T-2)\times 1}$$

此时是 T-2 期，因为需要考虑滞后一期。
因为

$$E(y_{is}\Delta u_{it})=E[y_{is}(u_{it}-u_{it-1})]=0,\forall s\leq t-2$$

因为 $E(y_{is}\Delta y_{it-1})\neq 0$
$y_{is}(s\leq t-2)$ 可以作为工具变量。
也可以使用向前和向后正交变换。

第八章 M 估计

8.1 GMM

工具变量个数大于内生变量个数时，使用降维，对所有外生变量回归进行降维。这种处理其实就是 GMM 的特例。

注 考试时会构造矩条件

思想：寻找样本矩条件最接近于 0 的参数。

例如 OLS$E(x'u) = 0$ 和 2sls$E(z'u) = 0$ 的假设。

当工具变量个数 (l) 大于内生变量个数 (k) 时, 上述方程对应着 l 个方程 k 个未知数, 此时, 最优解为最小化欧式距离 (Euclidean length) 的解 (向量的内积)

得到：

$$\arg\min_{\hat{\beta}} \left[\frac{1}{N}\sum_{i=1}^{N} z_i'(y_i - x_i\hat{\beta})\right]' \left[\frac{1}{N}\sum_{i=1}^{N} z_i'(y_i - x_i\hat{\beta})\right]$$

满足一致性，但不满足有效性。例如多的方程可能使得方程组无解。

为使 GMM 估计更有效, 可以引入权重矩阵 $\hat{W}$(也是一个估计量), 要求 $\hat{W}$ 必须是一个对称、半正定矩阵。此时 GMM 的参数估计式可以表达为:

其实是赋予一个权重。

此时得到的估计式为：

$$\arg\min_{\hat{\beta}} \left[\frac{1}{N}\sum_{i=1}^{N} z_i'(y_i - x_i\hat{\beta})\right]' \widehat{W} \left[\frac{1}{N}\sum_{i=1}^{N} z_i'(y_i - x_i\hat{\beta})\right]$$

不考虑样本容量:

$$\arg\min_{\hat{\beta}} \left[Z'(y - X\hat{\beta})\right]' \widehat{W} \left[Z'(y - X\hat{\beta})\right]$$

此时有最小值估计：

$$\hat{\beta} = \left(\mathrm{X'Z\hat{W}Z'X}\right)^{-1}\left(\mathrm{X'Z\hat{W}Z'Y}\right)$$

使用大数定理，可以得到

$$\begin{aligned}\hat{\beta} &= \left[\left(\frac{1}{N}X'Z\right)\hat{W}\left(\frac{1}{N}X'Z\right)'\right]^{-1}\left[\left(\frac{1}{N}X'Z\right)\hat{W}\left(\frac{1}{N}Z'y\right)\right]\\ &= \beta + \left[\left(\frac{1}{N}X'Z\right)\hat{W}\left(\frac{1}{N}X'Z\right)'\right]^{-1}\left[\left(\frac{1}{N}X'Z\right)\hat{W}\left(\frac{1}{N}Z'u\right)\right]\end{aligned}$$

由于 $E(z'u) = 0$, 所以 $\hat{\beta}$ 是 β 的一致性估计。

令 $C = E(X'Z)$, 假定 $\hat{W} \xrightarrow{p} W$

式子变形为三明治：

三明治都不是最有效的。

$$\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{p} (CWC')^{-1}CW\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(z_i'u_i)\right]+o_p(1)$$

$$\text{Avar}(\hat{\beta})=\frac{1}{N}(CWC')^{-1}CW\cdot\Lambda\cdot WC'(CWC')^{-1}$$

$$\Lambda\equiv E(z_i'u_iu_i'z_i)=\text{var}(z_i'u_i)=\frac{1}{N}\sum_{i=1}^{N}z_i'u_iu_i'z_i$$

u_i 是向量。

8.2 GMM 与 2SLS 的关系

有权重矩阵 $\widehat{W}=\left(N^{-1}\sum_{i=1}^{N}z_i'z_i\right)^{-1}=(Z'Z/N)^{-1}$

$$\begin{aligned}\hat{\beta}_{GMM}&=(X'Z(Z'Z/N)^{-1}Z'X)^{-1}(X'Z(Z'Z/N)^{-1}Z'y)\\&=(X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'y)\\&=(X^{*\prime}X)^{-1}(X^{*\prime}y)=\hat{\beta}_{2SLS}\end{aligned}$$

$$\text{Avar}(\hat{\beta})=\frac{1}{N}(C\Lambda^{-1}C')^{-1}$$

为了使得估计最优，取 $W=\Lambda^{-1}$ 进行消元。使得三明治变薄，此时渐进方差最小。
Λ 是矩条件的方差，是变量和残差的关系：
一般使用一阶段 GMM 估计：

$$\tilde{\beta}=\hat{\beta}_{\text{2SLS}}\to\widehat{W}=(Z'Z/N)^{-1}$$

$$\hat{\Lambda}=N^{-1}\sum_{i=1}^{N}z_i'\tilde{u}_i\tilde{u}_iz_i$$

加上这一步就是两阶段 GMM。

两阶段 GMM 的残差可以不断迭代进行估计。stata 命令 gmm 中需要选定参数：一阶段为 one step 二阶段为 two step 迭代参数是 igmm。

三阶段 SLS 是先估计 $\hat{u}\hat{u}'$, 同方差时最有效。

8.3 假设检验

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}z_i'u_i\right)'\widehat{W}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}z_i'u_i\right)\sim\chi_L^2$$

检验 Q 个线性约束：

约束函数和无约束函数进行比较：

假设无约束正确，则应当最小，一般为有约束减去无约束。

$$\left[\left(\sum_{i=1}^{N} z_i'\tilde{u}_i\right)' \hat{W} \left(\sum_{i=1}^{N} z_i'\tilde{u}_i\right) - \left(\sum_{i=1}^{N} z_i'\hat{u}_i\right)' \hat{W} \left(\sum_{i=1}^{N} z_i'\hat{u}_i\right)\right] / N \sim \chi_Q^2$$

过度识别检验：
原假设成立，服从卡方分布，残差替换误差项依旧服从，只是自由度减少。
过度

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i'\hat{u}_i\right)' \hat{W} \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i'\hat{u}_i\right) \sim \chi_{l-k}^2$$

Hansen J test 和 Sargan test 的区别
此时检验 p 越大越好，希望不通过

8.4 动态 GMM 面板

动态面板常常使用 GMM 估计，
动态完备性：控制当前时期后，前期和滞后期可以成为工具变量。
而组内去均值由于累计所有期，残差和滞后项相关，因此不能使用。
差分方式：
对于：

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}$$

矩条件就是：
此处，D 是一阶差分

$$E(Z_i' D u_i)_{L\times 1} = E\left(Z_i'\left(D y_i - \rho D y_{i(-1)}\right)\right) = 0$$

对于其渐进性：

$$\sqrt{N}(\hat{\rho}_{abond} - \rho) = \left(C'\widehat{W}C\right)^{-1} C'\widehat{W}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i' D u_i\right)$$

$$\operatorname{var}(\hat{\rho}_{abond}) = \frac{1}{N}\left(C'\widehat{W}C\right)^{-1} C'\widehat{W}\cdot\Lambda\cdot\widehat{W}C\left(C'\widehat{W}C\right)^{-1}$$

$$\Lambda \equiv E(Z_i' D u_i u_i' D' Z_i) = \operatorname{var}(Z_i' D u_i) = \frac{1}{N}\sum_{i=1}^{N} Z_i' D u_i u_i' D' Z_i$$

系统 GMM：
差分中预测前期 y 能力有限, 如果是随机游走，其实是弱工具变量。
$\rho < 1$ 时假设容易被满足。
加入假设 $E(x_{it}' c_i)$ 时不变：

$$E(\Delta x_{it}' c_i) = E(x_{it}' c_i) - E(x_{it-1}' c_i) = 0$$

$$E(\Delta x'_{is} v_{it}) = E(\Delta x'_{is} c_i) + E(x'_{is} u_{it}) - E(x'_{is-1} u_{it}) = 0(s \le t)$$

必要条件检验：一阶强相关，二阶不相关。排除误差存在序列相关。

推荐参考面板专题 | 差分 GMM 和系统 GMM 估计原理与 Stata 代码实现

注 这里考代码解读和矩阵列出来。重点注意工具变量有效性检验和工具变量的数量。

需要汇报 Hansen-J 统计量, 检验工具变量有效性 (OverIdentification Test)。

例如 2022 年的数据 $y_{2022} = y_{2021} + x_i + u$

其一阶滞后的水平差分工具变量为 y_{2020}

其一阶滞后的系统差分工具变量为 $y_{2020} - y_{2019}$。($y_{2021} - y_{2020}$) 是零阶差分。

如果在工具变量参数 iv(d1-d10) 中出现这种虚拟变量作为工具变量，因为为了避免共线性，会自动去掉一期，此时计数要减少一个。

Listing 8.1: Difference GMM

```
use housing_savingrate,clear

xtabond2

lnhdeposit l.lnhdeposit l.lnhp l.lnave_gdp

lnpop gender_ratio midscbiye_ratio primarybiye_ratio midschool_ratio primarysc_ratio
    kindergarton_ratio age0_14ratio age15_64ratio year_d1-year_d8,

gmm(l.lnhdeposit l.lnhp l.lnave_gdp, eq(diff) lag(2 .))

iv(lnpop gender_ratio midscbiye_ratio primarybiye_ratio midschool_ratio primarysc_ratio
    kindergarton_ratio age0_14ratio age15_64ratio,mz)

robust twostep nolevel
```

Listing 8.2: System GMM

```
use housing_savingrate,clear xtabond2

lnhdeposit l.lnhdeposit l.lnhp l.lnave_gdp

lnpop gender_ratio midscbiye_ratio primarybiye_ratio midschool_ratio primarysc_ratio
    kindergarton_ratio age0_14ratio age15_64ratio year_d1-year_d8,

gmm(l.lnhdeposit l.lnhp l.lnave_gdp, eq(diff) lag(2 .))

iv(lnpop gender_ratio midscbiye_ratio primarybiye_ratio midschool_ratio primarysc_ratio
    kindergarton_ratio age0_14ratio age15_64ratio,mz)

robust twostep nolevel
```

注 注意代码中的参数 lag(2 .) 就是二阶滞后到所有。如果是 lag(1 1) 就是一阶滞后到一阶滞后，也就是只要一阶滞后作为工具变量。

第九章 处理效应和 DID、DDD

9.1 处理效应

$D_i = 1$ 表示实验组和处理组；$D_i = 0$ 表示控制组。

Y_i^1 为个体 i 受政策影响时的结果 (outcome),Y_i^0 为不受政策影响时的结果。

实际观察到的数据：

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0 = Y_i + (Y_i^1 - Y_i^0) D_i$$

简单来说，当一个人已经被影响或者做出了选择，我们很难直接得到他如果没有被影响，或者做出另外一个选择时的结果。观察不到的是**反事实**。

i	D	Y	Y^1	Y^0	$Y^1 - Y^0$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

图 9.1: 潜变量的一个例子

平均处理效应：

$$ATE = E\left(Y_i^1 - Y_i^0\right) = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i^1 - Y_i^0\right)$$

处理组平均效应：

$$ATT = E\left(Y_i^1 - Y_i^0 \mid D_i = 1\right) = \frac{1}{N_1}\sum_{i=1}^{N_1}\left(Y_i^1 - Y_i^0\right)$$

控制组平均处理效应：

$$ATU = E\left(Y_i^1 - Y_i^0 \mid D_i = 0\right) = \frac{1}{N_0}\sum_{i=1}^{N_0}\left(Y_i^1 - Y_i^0\right)$$

设置 $Y_i^0 = \mu_i^0 + v_i^0$, $\quad Y_i^1 = \mu_i^1 + v_i^1$,

即可得到：

$$Y_i = \mu_i^0 + \left(\mu_i^1 - \mu_i^0\right) D_i + D_i \left(v_i^1 - v_i^0\right) + v_i^0 \tag{9.1}$$

其实是 $Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$ 的展开，因为我们只能看到最终的处理结果，最终呈现的就是处理和非处理的期望和。

对应的**条件期望**为

$$E(Y_i|D_i)=\mu_i^0+\left[E\left(\mu_i^1-\mu_i^0|D_i\right)+E\left(v_i^1-v_i^0|D_i\right)\right]*D_i+E\left(v_i^0|D_i\right)$$

接下来进一步分析各个效应的对应条件。

对于 ATE 来说：

$$ATE=\left(\mu_i^1-\mu_i^0\right)$$

$$Y_i=\mu_i^0+\beta^{ATE}D_i+u_i$$

其中 $u_i=D_i\left(v_i^1-v_i^0\right)+v_i^0$

通过 ols 部分可以知道，ATE 无偏估计要求 $\mathrm{Cov}(D_i,v_i^0)=0,\mathrm{Cov}(D_i,v_i^1-v_i^0)=0$

对于 ATT 来说，

$$ATT=E\left(\mu_i^1-\mu_i^0|D_i=1\right)+E\left(v_i^1-v_i^0|D_i=1\right)$$

代入式子 (9.1) 可以得到：

$$E(Y_i|D_i)=\mu_i^0+\left[E\left(\mu_i^1-\mu_i^0|D_i\right)+E\left(v_i^1-v_i^0|D_i\right)\right]*D_i+E\left(v_i^0|D_i\right)$$

那么：

$$\begin{aligned}&E(Y_i|D_i=1)-E(Y_i|D_i=0)\\&=ATT+E\left(v_i^0|D_i=1\right)-E\left(v_i^0|D_i=0\right)\end{aligned}$$

因此其无偏估计的条件假设为 $E(v_i^0|D_i)=0$。

练习 9.1 要求 ATE 和 ATT 谁要求更严格？

解 [要求] 解答

ATE 更严格，同时要求 $E(v_i^0|D_i)=0,E(v_i^1-v_i^0|D_i)=0$

ATT 只要求 $E(v_i^0|D_i)=0$

注 这里可以看出 ATT 的假定弱于 ATE 的假定，因为 ATT 只保证了局部的严格情况，而 ATE 需要整体都严格。

练习 9.2 要求 如何理解 ATE 无偏的两个要求。$E(v_i^0|D_i)=0,E(v_i^1-v_i^0|D_i)=0$

解 [ATE 无偏的两个要求理解] 如下

(1) 选择偏误：政策是随机的, 政策的实施没有因为个体先天的差异 v_i^0 而不同;

(2) 排序偏误：个体参与政策也是随机的, 个体参与政策也没有因为私人收益 $v_i^1-v_i^0$ 的不同而不同。

如果加入控制变量，

$$Y_i=\mu_i^0+\beta^{ATE}D_i+X_i\gamma+u_i$$

此时无偏需要的假定为：

$$E\left(v_i^0|D_i\right)=0;E\left[v_i^1-v_i^0|D_i\right]=0$$

变成了

$$Y_i^1,Y_i^0\perp D_i\mid X_i$$

此时有

$$E\left(Y^1|X,D\right)=E\left(Y^1|X\right);E\left(Y^0|X,D\right)=E\left(Y^0|X\right)$$

此时 ATT=ATE，因为变形可得到

$$\begin{aligned}ATT &= E\left(Y_i^1-Y_i^0 \mid D_i=1,X\right)\\ &= E\left(Y_i \quad |D_i=1,X\right)-E\left(Y_i \quad |D_i=0,X\right)\\ &= E\left(Y_i^1 \mid X\right)-E\left(Y_i^0 \mid X\right)=ATE\end{aligned}$$

因此这个假定被叫做 ATE 假定 1(ignorability of treatment) 或 Unconfoundedness(无混杂因子)。

ATE 假定 2 为

$$0<pr(D_i=1 \mid X_i)<1$$

直观的含义是能同时找到对照组和实验。违背的情况下，例如只能找到处理组 $pr(D_i=1 \mid X_i)=1$，自然对照组的概率为 0。

那么 IV 需要 Z_i 和 D_i 相关，却要求 Z_i 和 $(v_i^1-v_i^0)D_i$ 不相关。所以这时候寻找 IV 是非常苦困难的，需要 D_i 和 $v_i^1-v_i^0$ 相互独立，此时可以证明 $Cov(Z_i,(v_i^1-v_i^0)D_i)=0$。也就是不存在排序偏误时比较容易找 IV。

反事实（考过描述反事实）：$Y_i^0|D_i=1, Y_i^1|D_i=0$

9.2 DID

$$Y_{it}^0=\mu_i^0+c_i+\rho\varphi_t+\varepsilon_{it}^0$$

$$Y_{it}^1=\mu_i^1+c_i+\rho\varphi_t+\varepsilon_{it}^1$$

假定 t-1 期为政策前,t 期为政策后, 此时, 可以观测到的数据为:

$$Y_{it-1}=Y_{it-1}^0=\mu_i^0+c_i+\rho\varphi_{t-1}+\varepsilon_{it-1}^0$$

$$Y_{it}=\mu_i^0+\left(\mu_i^1-\mu_i^0\right)D_{it}+\left(\varepsilon_{it}^1-\varepsilon_{it}^0\right)D_{it}+c_i+\rho\varphi_t+\varepsilon_{it}^0=\alpha+\beta^{ATT}D_{it}+c_i+\rho\varphi_t+u_{it}$$

$$\Delta Y_{it}=\beta^{ATT}\Delta D_{it}+\rho\Delta\varphi_t+\Delta u_{it}$$

$$\Delta D_{it}=D_i\times 1-D_i\times 0=D_i$$

$$\begin{aligned}\beta^{\text{ATT}} &= E(\Delta Y_{it} \mid D_i=1)-E(\Delta Y_{it} \mid D_i=0)\\ &= E(Y_{it} \mid D_i=1)-E(Y_{it-1} \mid D_i=1)-[E(Y_{it} \mid D_i=0)-E(Y_{it-1} \mid D_i=0)]\\ &= \underbrace{E(Y_{it}^1 \mid D_i=1)-E(Y_{it-1}^0 \mid D_i=1)}_{\text{first difference}}-\underbrace{\left[E(Y_{it}^0 \mid D_i=0)-E(Y_{it-1}^0 \mid D_i=0)\right]}_{\text{second difference}}\end{aligned}$$

可以看到我们目前是用的对照组的政策前后的差距作为反事实，目前连 ATT 都算不上。ATT 需要 $E(Y_{it}^1 - Y_{it}^0|D_i = 1)$。$D_i$ 表示对照组和实验组 =1 为接受冲击。=0 代表未接受冲击。Y_{it}^0 为实际数据没有经受过冲击。

重点——平行趋势检验

注[关于平行趋势检验]

$$E\left(Y_{it}^0|D_i = 0\right) - E\left(Y_{it-1}^0|D_i = 0\right) = E\left(Y_{it}^0|D_i = 1\right) - E\left(Y_{it-1}^0|D_i = 1\right)$$

$E\left(Y_{it}^0|D_i = 1\right)$ 是反事实, 无法观测, 所以此充分条件无法检验。该条件的必要条件是政策发生前实验组和控制组具有共同趋势。平行趋势检验是必要条件。

考过（不满足平行趋势的情况）：也就是说实验组的平行趋势是通过对照组判断的，如果实验组在处理之前是**上偏**的，处理效应实际上会被低估。

充分条件是事后的反事实 $(Y_{it}^0|D_i = 1)$（假如 t 期的实验组没有受到政策冲击），而唯一能检验的是事前的反事实，因**此平行趋势检验的是必要条件**。

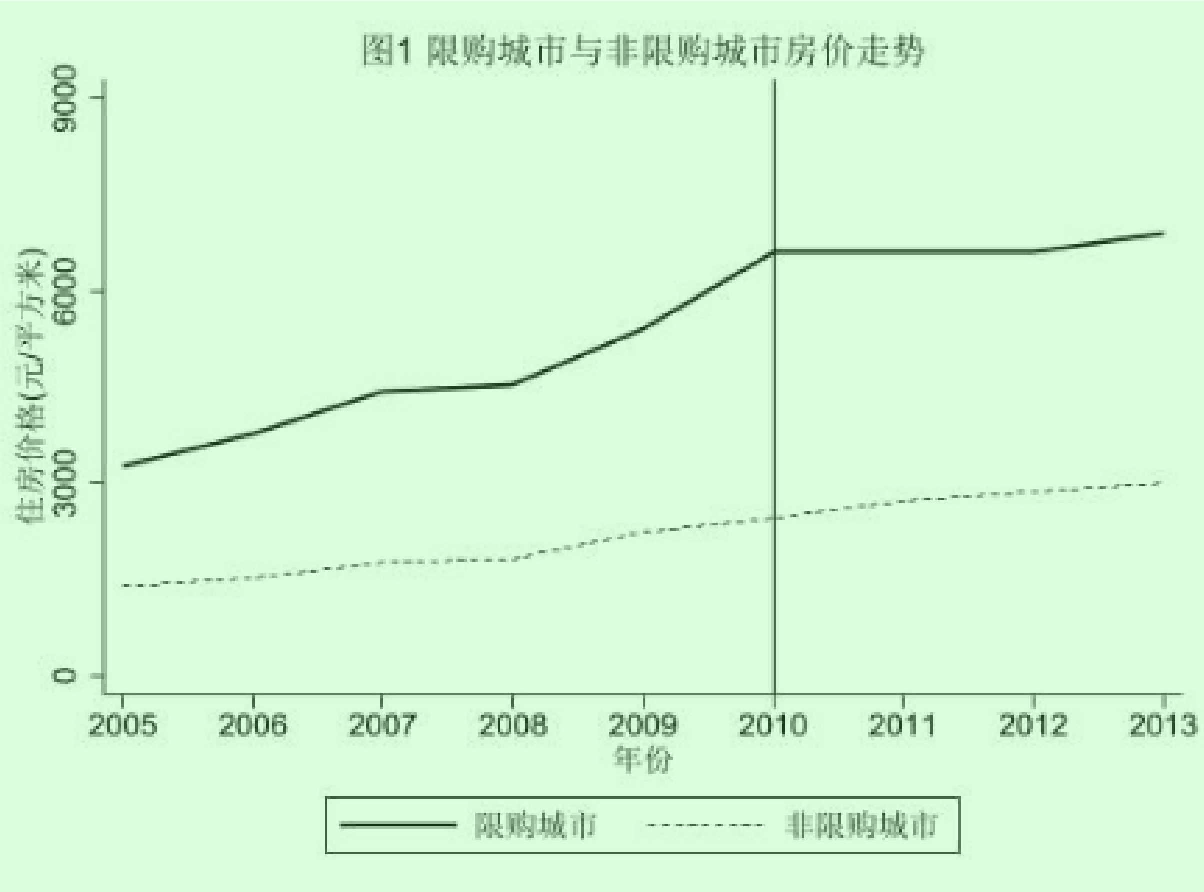


图 9.2: 上偏不满足平行趋势检验的例子

此时也是通过一阶差分消去了个体效应。

采用组内估计，意味着对于：

$$Y_{it} = \alpha + \beta^{ATT} D_{it} + X_{it}\gamma + c_i + \rho\varphi_t + u_{it}$$

这给出了更强的假设：

$$E\left[\varepsilon_{it}^{0}|D_i\right]=0 \Rightarrow E\left[\varepsilon_{it}^{0}-\varepsilon_{it-1}^{0}|D_{it}\right]=0 \Rightarrow E\left[\varepsilon_{it}^{0}-\bar{\varepsilon}^{0}|D_{it}\right]=0$$

第二个箭头: 相邻两期满足平行趋势　任何两期都满足。

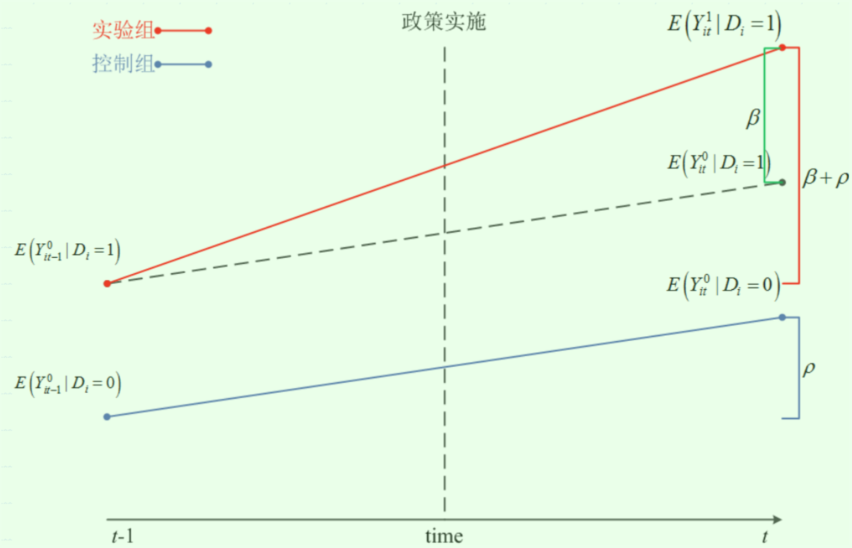


图 9.3: DID 图形演示

问题 9.1 式子:

$$ln(hp)_{ct}=\alpha_0+\alpha_1 xiangou_c\times post+\alpha_2 xiangou_c+\alpha_3 post+u_{ct}$$

ln(hp)ct 表示城市 c 在第 t 年房价的对数;xiangouc 表示城市 c 是否实施住房限购, 是则取值为 1, 否则取值为 0, 前者为实验组, 后者为控制组。

请解释 α1、α2、α3 的经济含义。

```
Linear regression                               Number of obs =     3171
                                                F(  3,    364) = 1215.57
                                                Prob > F       =   0.0000
                                                R-squared      =   0.5235
                                                Root MSE       =   .35556

                                (Std. Err. adjusted for 365 clusters in city_new)
------------------------------------------------------------------------------
             |               Robust
       ln(hp)|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  xiangou×post|  -.0295327   .0172272    -1.71   0.087    -.0634099    .0043446
      xiangou|   .8659204     .06244    13.87   0.000     .7431321    .9887088
         post|   .4824822   .0094772    50.91   0.000     .4638453    .5011192
        _cons|    7.39217   .0172621   428.23   0.000     7.358224    7.426116
------------------------------------------------------------------------------
```

图 9.4: 回归结果

解

注意回归式子的结构 (解释变量和被解释变量是否取了对数), 这会影响结果解读的百分比放缩解释。

α1 代表限购政策在实验组的平均处理效应是降低 2.9% 的房价.

α2 代表限购前, 实验组城市房价比对照组房价平均要贵 86.59%.

(解析: 实际上是通过 $(Y|post=0, xiangou_c=1)-(Y|post=0, xiangou_c=0)$)

α3 代表限购前后, 对照组房价平均上涨 48.25%.

(解析: 实际上是通过 $(Y|post=1, xiangou_c=0)-(Y|post=0, xiangou_c=0)$)

这也是理解虚拟变量交互项经济含义的过程!!

注 考过，只有两期的面板，控制低次项作用和个体时间固定效应效果相同。证明。

9.3 CIC(略)

传统 DID 只能处理可加 (additive) 变量的函数形式。

Athey & Imbens(2006, Econometrica)提出了一种基于函数分布来估计处理效应的一般化方法—双重变换法 (Changein-Change):

- DID 是 CIC 的一个特例;
- 可以处理非线性 DID;
- 可以估计分位数 (因变量分位数) 处理效应 (Quantile DID)。

CIC 放松了 DID 估计中需要满足平行趋势的前提假定。但取而代之需要满足如下假定：

1. **CIC 假设条件 1**：未受政策影响的潜在结果与观察不到的异质性之间是单调函数，即 $Y_i^0=h(U_i,T_i)$ 是 U 的单调函数。
2. **CIC 假设条件 2**：未观察到的异质性的分布在组内 (实验组/控制组) 随着时间的推移保持不变，即 $U_i \perp T_i \mid D_i$。允许未观察到的异质性的分布在组间有差异。

上述两条假设下：

实验组 Y_i^0 的分布和控制组 Y_i^0 的分布具有一一对应的关系 (尽管二者的分布可能不同)

实验组 Y_i^0 和控制组 Y_i^0 的分布是时不变的 (同时也是一一对应的)。

CIC 的假设条件可以总结为：**分布平衡趋势假定**

定义 9.1 (cic)

定义 $Y^p_{D_i,t}(D_i=0,1;t=0,1;p=0,1)$ 为组 D_i 在时间 $t(t=0$ 表示政策前，$t=1$ 表示政策后) 下潜在状态为 p 的结果变量，则 CIC 估计量为：

$$\tau^{\text{CIC}}=\mathbb{E}\left[Y_{11}^1-Y_{11}^0\right]=\mathbb{E}\left[Y_{11}^1\right]-\mathbb{E}\left[F_{Y,01}^{-1}\left(F_{Y,00}(Y_{10})\right)\right]$$

♣

Y_{11}^1 的反事实 Y_11^0 是通过分布反推出来的：

1. 给定实验组 $t=0$ 时的值 $Y_10\Rightarrow$
2. 寻找其在控制组 t =0 时期的分位数 $F_Y,00(Y_{10})\Rightarrow$
3. 由于 $F_Y,00(y)$ 和 $F_Y,01(y)$ 的分布时不变，因此，可以在 $t=1$ 期控制
4. 组中去找对应分位数 $F_Y,00(Y_{10})$ 的分位数点 $F_Y,01^{-1}\left(F_{Y,00}(Y_{10})\right)\Rightarrow$
5. 在分布平衡趋势假定下，该分位数点也就是 Y_11^0。

例题 9.1cic

考虑一个 Probit 模型

$$Y_{it}=\mathbf{1}\{\alpha+\beta D_{it}+c_i+\rho\varphi_t+u_{it}>0\}$$

其中，u_it 服从标准正态分布 $(u_it\sim N(0,1))\mathbf{1}\{\cdot\}$ 是指示函数。此时有

$$E\left(Y_{it}|D_{it},c_i,\varphi_t\right)=\Phi\left(\alpha+\beta D_{it}+c_i+\rho\varphi_t\right)\Phi^{-1}\left[E\left(Y_{it}|D_{it},c_i,\varphi_t\right)\right]=\alpha+\beta D_{it}+c_i+\rho\varphi_t$$

可以得到处理效应系数 β 的表达式为：

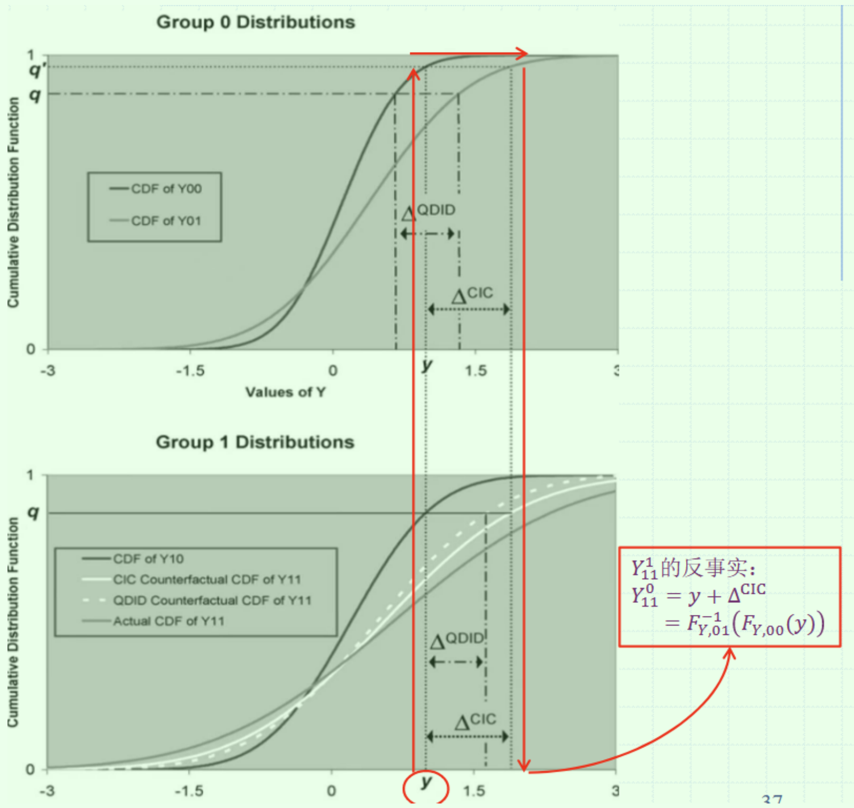


图 9.5: CIC 图示

$$
\begin{aligned}
&= \{\Phi^{-1}[E(Y_{it}|D_i=1)] - \Phi^{-1}[E(Y_{it-1}|D_i=1)]\} \\
&-\{\Phi^{-1}[E(Y_{it}|D_i=0)] - \Phi^{-1}[E(Y_{it-1}|D_i=0)]\} \\
&= \left\{\Phi^{-1}\left[E\left(Y_{it}^1|D_i=1\right)\right] - \Phi^{-1}\left[E\left(Y_{it-1}^0|D_i=1\right)\right]\right\} \\
&- \left\{\Phi^{-1}\left[E\left(Y_{it}^0|D_i=0\right)\right] - \Phi^{-1}\left[E\left(Y_{it-1}^0|D_i=0\right)\right]\right\}
\end{aligned}
$$

其中，Y_{it}^1 是在受到处理 $D_{it}=1$ 时的潜在结果，Y_{it}^0 是在未受处理 $D_{it}=0$ 时的潜在结果。在分布平衡趋势假定下有:

$$
\Phi^{-1}\left[E\left(Y_{it}^0|D_i=0\right)\right] - \Phi^{-1}\left[E\left(Y_{it-1}^0 \mid D_i=0\right)\right] = \Phi^{-1}\left[E\left(Y_{it}^0|D_i=1\right)\right] - \Phi^{-1}\left[E\left(Y_{it-1}^0 \mid D_i=1\right)\right]
$$

那么可得:

$$
\beta = \Phi^{-1}\left[E\left(Y_{it}^1 \mid D_i=1\right)\right] - \Phi^{-1}\left[E\left(Y_{it}^0 \mid D_i=1\right)\right]
$$

于是有

$$
E\left(Y_{it}^0 \mid D_i=1\right) = \Phi\left[\Phi^{-1}\left(\bar{Y}_{it}^{11}\right) - \beta\right]
$$

因此得到

$$
ATT = E\left(Y_{it}^1 \mid D_i=1\right) - E\left(Y_{it}^0 \mid D_i=1\right) = \bar{Y}_{it}^{11} - \Phi\left[\Phi^{-1}\left(\bar{Y}_{it}^{11}\right) - \beta\right]
$$

CIC 范文:Adrienne M. Lucas & Isaac M. Mbiti, 2012. “Access, Sorting, and Achievement: The Short-Run

Effects of Free Primary Education in Kenya, ” American Economic Journal: Applied Economics, 4(4): 226-253.

注[cic 使用] logit 类型的模型不能用平行趋势，因为概率难以减掉，这时候考虑分布的检验

更进一步的，平行趋势与 y 函数的关系——Y or log(Y)—平行趋势何时对函数设定形式敏感?

9.4 DDD

DDD 只能解决意向处理效应的异质性。

换而言之，就是双重查分的划分不够严谨，再次加一个条件进行对照。

其估计：

$$\begin{aligned}
& DDD \\
=& \left\{\left[E\left(Y_{it}^{1}|D_i=1,Q_i=1\right)-E\left(Y_{it-1}^{0}|D_i=1,Q_i=1\right)\right]\right. \\
&-\left.\left[E\left(Y_{it}^{0}|D_i=1,Q_i=0\right)-E\left(Y_{it-1}^{0}|D_i=1,Q_i=0\right)\right]\right\} \\
&-\left\{\left[E\left(Y_{it}^{0}|D_i=0,Q_i=1\right)-E\left(Y_{it-1}^{0}|D_i=0,Q_i=1\right)\right]\right. \\
&-\left.\left[E\left(Y_{it}^{0}|D_i=0,Q_i=0\right)-E\left(Y_{it-1}^{0}|D_i=0,Q_i=0\right)\right]\right\}
\end{aligned}$$

其假设条件（平行趋势）

$$\begin{aligned}
&\left[E\left(Y_{it}^{0}|D_i=0,Q_i=1\right)-E\left(Y_{it-1}^{0}|D_i=0,Q_i=1\right)\right] \\
&-\left[E\left(Y_{it}^{0}|D_i=0,Q_i=0\right)-E\left(Y_{it-1}^{0}|D_i=0,Q_i=0\right)\right] \\
&=\left[E\left(Y_{it}^{0}|D_i=1,Q_i=1\right)-E\left(Y_{it-1}^{0}|D_i=1,Q_i=1\right)\right] \\
&-\left[E\left(Y_{it}^{0}|D_i=1,Q_i=0\right)-E\left(Y_{it-1}^{0}|D_i=1,Q_i=0\right)\right]
\end{aligned}$$

例子：连享会

注 考过：三重差分中，表示出低次项系数，并解释经济含义。

推荐参考：三重差分模型中的二重交互项的系数都有意义吗？- 袁迎的回答

核心思路就是控制变量减去消元，解释经济含义。也就是计量的可比性。

正因如此，计量差分一般最多三重差分，再上面难以从直觉上理解和解释。