计量经济学笔记

ElegantLaTeX 版本

作者：滑翔闪

组织：ElegantLaTeX Program

时间：September 17, 2025

ElegantLaTeX Program

因果推断是一门哲学

目录

第一章 概率论与数理统计预备知识：渐近理论 **3**

1.1 数列的收敛 . . . 3
1.2 随机向量或矩阵的收敛 . . . 3
1.3 随机向量或矩阵收敛的一些性质 . . . 4
1.4 随机向量分布收敛的性质 . . . 5
1.5 随机向量分布收敛的性质 . . . 5
1.6 渐进性 . . . 6
1.7 线性代数相关知识 . . . 6

第二章 简单 OLS **7**

2.1 研究对象 . . . 7
2.2 希望性质 . . . 8
2.3 最小二乘 . . . 9
2.3.1 相关性分析到最小二乘 . . . 9
2.3.2 均方误最小 . . . 11
2.3.3 拟合优度 . . . 13
2.3.4 矩估计 . . . 14
2.3.4.1 期望 . . . 15
2.3.4.2 方差 . . . 17

第三章 多变量 OLS **22**

3.1 Frisch-Waugh 定理 . . . 22
3.2 遗漏变量 . . . 28
3.3 控制无关变量 . . . 28
3.4 系数含义 . . . 29
3.5 匹配 . . . 29
3.6 多重共线性 . . . 30

第四章 渐进分布假设检验 **31**

4.1 分布基础知识 . . . 31
4.2 推断统计假设 . . . 32
4.3 渐进分布 . . . 33
4.4 OLS 条件估计总结 . . . 36

4.5 假设检验 37
4.6 t 检验 39
4.7 F 检验 39
4.8 渐进检验 41

第五章 工具变量 **43**
5.1 内生性与外生性 43
5.2 工具变量 43
5.2.1 期望 43
5.2.2 方差 46
5.3 2SLS 47
5.4 工具变量的选择 47
5.5 假设检验 49
5.5.1 内生性 Hausman 检验 49
5.5.2 弱工具变量 49
5.5.3 过度识别约束变量 49
5.6 LATE 51

第六章 现象面板固定随机效应 **53**
6.1 面板数据 53
6.2 假设 53
6.3 随机效应 53
6.4 固定效应 55
6.5 比较 58
6.6 检验 60
6.6.1 思路 1 60
6.6.2 思路 2 60
6.6.3 思路 3 61

第七章 面板内生性 **62**
7.1 内生性来源 62
7.2 工具变量 62
7.2.1 随机效应 62
7.2.2 固定效应 62
7.3 序列外生的动态面板 63
7.3.1 FD 64

第八章 M 估计 **66**

8.1 GMM 66

8.2 GMM 与 2SLS 的关系 67

8.3 假设检验 68

8.4 动态 GMM 面板 68

第九章 处理效应和 DID、DDD **71**

9.1 处理效应 71

9.2 DID 73

9.3 DDD 75

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

$$plim g(x_N) = g(plim x_N) = g(c)$$

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

$$\forall K\times 1 维非随机向量 c 满足 c'c=1, c'x_N \rightarrow c'x$$

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)

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

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

假如随机向量 $\boldsymbol{x}_N \xrightarrow{d} \boldsymbol{x}$, $g{:}R^K \rightarrow R^J$ 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 $\lim_{N\rightarrow\infty} F_N(x) = F(x)$ (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 $R_{Q\times P}, Rank(R) = Q$,

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

Delta method：

推荐参考义——Delta method

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

定义 $C(\theta) = \nabla_\theta c(\theta)$ 为函数 的 × 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 β_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。

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 2.1 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 2.2 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。

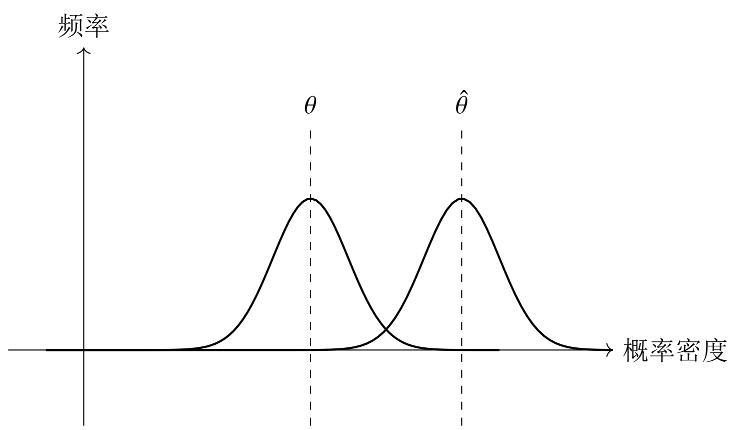
标准误是标准差的估计量。

置信区间：重复抽样,n% 的区间包含确定值。

2.2 希望性质

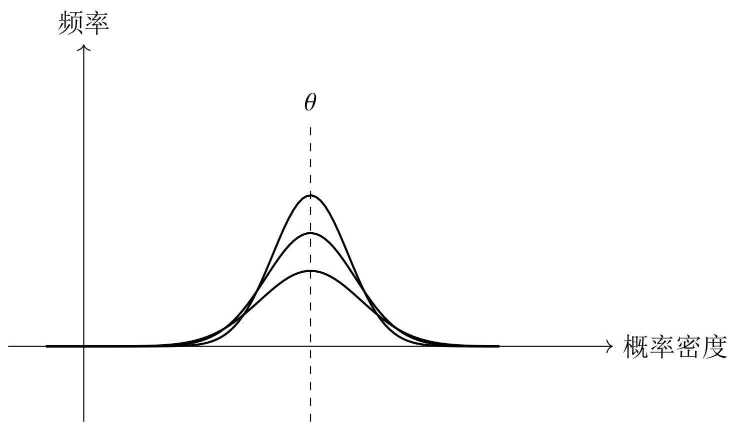
无偏性：重复抽样, 相同数量, 不同样本。

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。

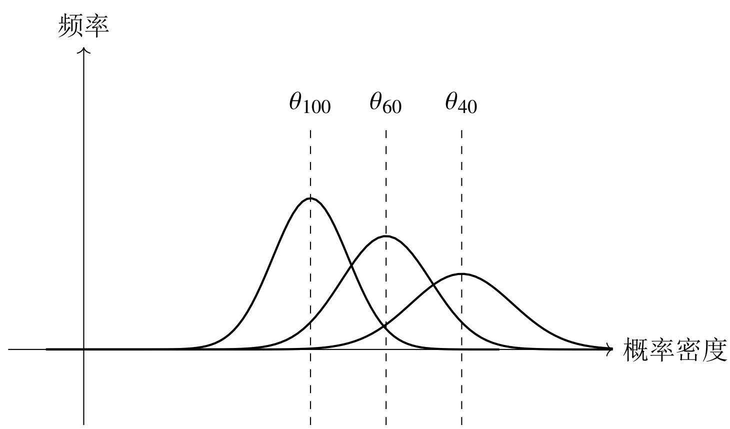


有效性：不同方法, 相同样本, 相同数量, 方差更小。

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



命题 2.1

一致性和无偏性无关。

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）

例题 2.3 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 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。

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

。

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）可证明 r_{xy} 范围在-1 到 1 之间

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。

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

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 $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$。

借助以上式子，最小二乘的系数估计就可以变为：

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

命题 2.2 (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。

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

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

$$SST = SSR + SSE$$

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

$$RMSE = \sqrt{MSE}$$

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

2.3.4 矩估计

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)。

命题 2.3 (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）。
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. 同方差性
6. 无自相关
7. 正态性

   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）

外生性和内生性的讨论

对于外生性有以下性质：

外生性：$E(u) = E[E(u|X)] = 0$

严格外生性：$COV(u, x) = 0, E(uX) = 0$

迭代期望定律 (条件期望的期望等于无条件期望) 对 x 积分：$E(E(u|x)) = E(u)$

强外生性：误差和其他函数不相关。

$$\begin{aligned} cov(u, f(x)) = 0 \\ &= E(u, f(x)) - E(u)f(x) \\ &= E[E(u, f(x)|x)] \\ &= E[(u, f(x)|x)] = 0 \end{aligned}$$

OLS 的统计思想为 $minSSR(\beta_t) = \sum_{t=1}^{n}(y_t - \beta_t x_t)$
得到一阶条件：$\frac{\partial SSR}{\partial \beta_t} = 0$
转换成矩阵形式：

$$\begin{aligned} SSR &= (Y_i - X_i\hat{\beta}_i)(Y_i - X_i\hat{\beta}_i)' \\ &= Y_iY_i' - X_iY_i'\hat{\beta}_i - \hat{\beta}_i'X_i'iY_i + \hat{\beta}_iX_i'X_i\hat{\beta}_i \end{aligned}$$

研究一阶条件解得

$$\hat{\beta}_i = (X_i'X_i)^{-1}X_i'Y_i = X_i^{-1}Y_i(\text{简化形式})$$

简化条件要求 X 是方阵可求逆, 如果并非方阵, 则要求 $n = k + 1$, 但这种情况在现实几乎不存在。

参数的矩阵求解为：

$$\beta = \left[\mathrm{E}(\mathbf{x}'\mathbf{x})\right]^{-1}\mathrm{E}(\mathbf{x}'\mathbf{y})$$

2.3.4.1 期望

矩方法就是使用样本均值替换估计量进行估计，于是得到:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right]^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{y}_i\right] \\ &= \boldsymbol{\beta} + \left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{x}_i\right]^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'u_i\right] \end{aligned}$$

引理 2.1 (OLS 一阶条件)

观察残差定义可以发现：

$$\hat{y} = X\hat{\beta}$$

则

$$\hat{\mu} = y - \hat{y} = y - X\hat{\beta}$$

而一阶条件可以进行转化：

$$\begin{aligned}&\frac{\partial SSR}{\partial \beta_t}\\&=X_i'Y_i-X_i'X_i\hat{\beta}_i\\&=X_i(Y_i-X_i\hat{\beta}_i)\\&=X'\hat{u}=0\end{aligned}$$

因此一阶条件意味着只要有截距，残差平方和即为 0。也就是说 $\hat{\beta}$ 是 β 的有条件无偏估计。

此处得带求解的 **OLS 假设 1**：$E(x'u)-0$

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 2.3 (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：

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\hat{\beta}_2x_2$$

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：

$$\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1x_1$$

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

$$x_2 = \tilde{\delta_1} x_1 + u$$

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。

2.3.4.2 方差

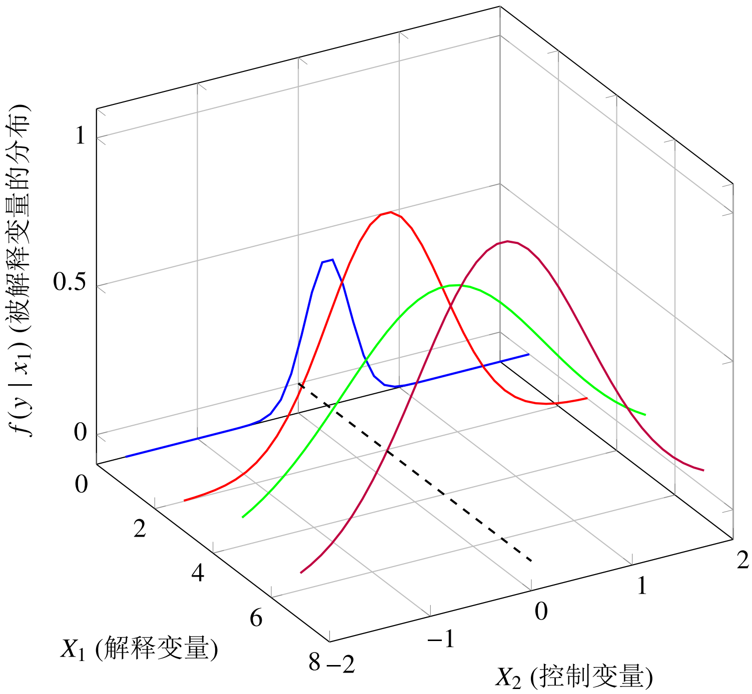
假定 5. 同方差性和不存在序列自相关
是为了获取 $\hat{\beta}$ 最简单的协方差矩阵。
(i)$\mathrm{Var}(u_i|\mathbf{X}) = \sigma^2; t = 1, 2, \cdots; n$。
同方差假定,u_t 的方差不依赖于 X 中的任意一个元素, 且不同观测 t 的方差都相等。
异方差示意图,
由于 $var(u|x) = var(y|x)$, 因此只要 $var(u|x)$ 是 x 的函数, 就会出现**异方差**。
证明 证明 $var(u|x) = var(y|x)$
首先

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

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 2.4 (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。

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

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

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

定义 2.2

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

此时估计量是最优线性无偏估计。

这种假定可以如此总结：

假定 1-4：$E(Y|X) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$

假定 5：$Var(Y|X) = \sigma^2$

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。 ♣

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。

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

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 2.5

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)。

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

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)。

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”。

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：

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

$$-2\sum_{i=1}^{n} x_{ij}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}\right) = 0, \forall j = 1, \cdots, k$$

接下来对 x_{i1} 进行回归分解 $x_{i1} = \hat{x}_{i1} + \hat{r}_{i1}$。代入式子，得到：

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

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

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 $\ddot{X}_2$,

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

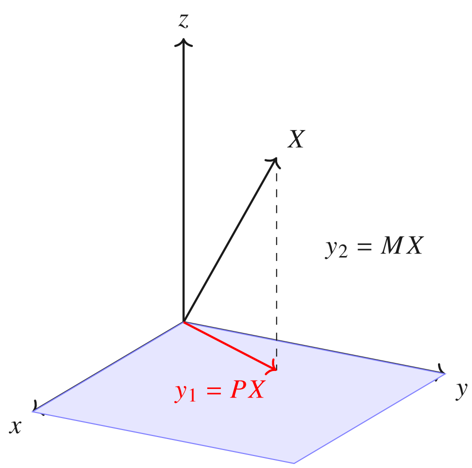
并在这个过程中得到投影矩阵 $P \equiv X(X'X)^{-1}X'$ 和消失矩阵 $M \equiv I - P$。

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

$$X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 = X_1'Y \tag{3.1}$$

$$X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 = X_2'Y \tag{3.2}$$

解第方程3.1得到

$$\hat{\beta}_1 = (X_1'X_1)^{-1}(X_1'Y - X_1'X_2\hat{\beta}_2) = (X_1'X_1)^{-1}X_1'(Y - X_2\hat{\beta}_2) \tag{3.3}$$

将方程3.3代入方程3.2, 得到

$$X_2'X_1(X_1'X_1)^{-'}X_1'(Y - X_2\widehat{\beta}_2) + X_2'X_2\widehat{\beta}_2' = X_2'Y \tag{3.4}$$

对于方程3.4, 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:

$$\ddot{y} = M_1 y = M_1(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{u})$$

基于消失矩阵的性质 $M_1\hat{\beta}_1 X_1 = 0$
得到

$$\ddot{y} = M_1 y = M_1 X_2 \hat{\beta}_2 + M_1 \hat{u}$$

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

$$M_1 y = M_1 X_2 \hat{\beta}_2 + \hat{u}$$

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 $X_2 = \ddot{\beta}_2 X_1 + \ddot{X}_2$,
得到残差 $\ddot{X}_2$,

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

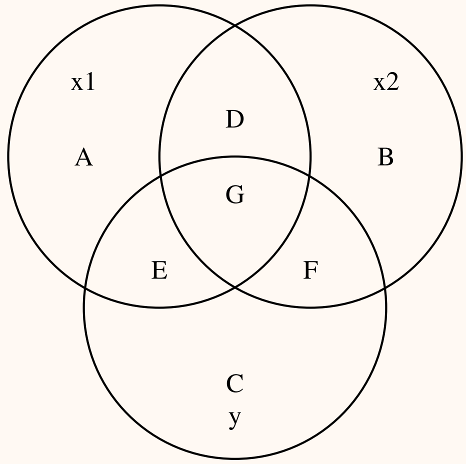
在上式中, 如果 $X_1'X_2 = 0$, 可得 $\ddot{X}_2 = M_1 X_2 = X_2$
此时短回归式子 $y = \beta x_2 + u$ 中,

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

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。

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

接下来看其条件方差

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

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 遗漏变量

在遗漏了 x_2 的情况下

$$\hat{\beta}_1 = \beta_1 + \beta_2(x_1'Mx_1)^{-1}(x_1'Mx_2) + \beta_2(x_1'Mx_1)^{-1}(x_1'Mu)$$

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。

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 3.1 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。

3.3 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 $E(u|x_1, x_2) = 0$

$$\hat{\beta}_{b1} = \beta_1 + \beta_2(x_1'Mx_1)^{-1}(x_1'Mu)$$

其方差为 $var(\hat{\beta}_{b1}|x) = \sigma^2(x_1'Mx_1)^{-1}$

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。

3.4 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%

$$y = \beta lnx + u$$

x 变化 1%,y 变化 $\frac{\beta}{100}$

$$lny = \beta x + u$$

x 变化 1 单位,y 变化 100β%

只改变截距，不改变系数。

$$\log(y) = \alpha + \beta x + u \tag{3.5}$$

$$\log(c) + \log(y) = \log(c) + \alpha + \beta x + u \tag{3.6}$$

$$\log(cy) = [\log(c) + \alpha] + \beta x + u \tag{3.7}$$

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

3.5 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) 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。

3.6 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

。

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 $\eta \sim N(0, 1)$。

注 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。

$Y_j \sim N(\mu, \sigma^2)$ 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 $X \sim \chi_n^2$

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 $Y + Z \sim \chi_n + m^2$

***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。

F 分布

令 $X_1 \sim \chi^2_{k_1}$, $X_2 \sim \chi^2_{k_2}$, 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 $F \sim F_{k_1,k_2}$。

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 4.1

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。

定理 4.2

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2.3.4.2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 4.3

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 4.4 (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 4.5

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：

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

$o_p(1)$ 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 4.6 (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。

$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$ 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=$\sigma^2\mathbf{A}$。(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,$\mathbb{E}(\mathbf{x}_i' u_i u_i \dot{\mathbf{x}}_i) = \mathbb{E}[\mathbb{E}(u_i u_i|\mathbf{x}_i'\mathbf{x}_i)\mathbf{x}_i'\mathbf{x}_i] = \mathbb{E}[\mathbb{E}(u_i u_i|\mathbf{x}_i'\mathbf{x}_i)\mathbf{x}_i'\mathbf{x}_i] = \mathbb{E}[0 \cdot \mathbf{x},'\mathbf{x},] = 0$。协方差为零便意味着和的方差等于方差之和。但 $\mathrm{Var}[\mathbf{x}_i' u_i] = \mathbb{E}(\mathbf{x}_i' u_i u_i \mathbf{x}_i) = \mathbb{E}(u_i^2 \mathbf{x}_i'\mathbf{x}_i)$。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= $\sigma^2\mathbf{A}$, 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$(0, \sigma^2)$] 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

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

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)。

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)

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

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

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。

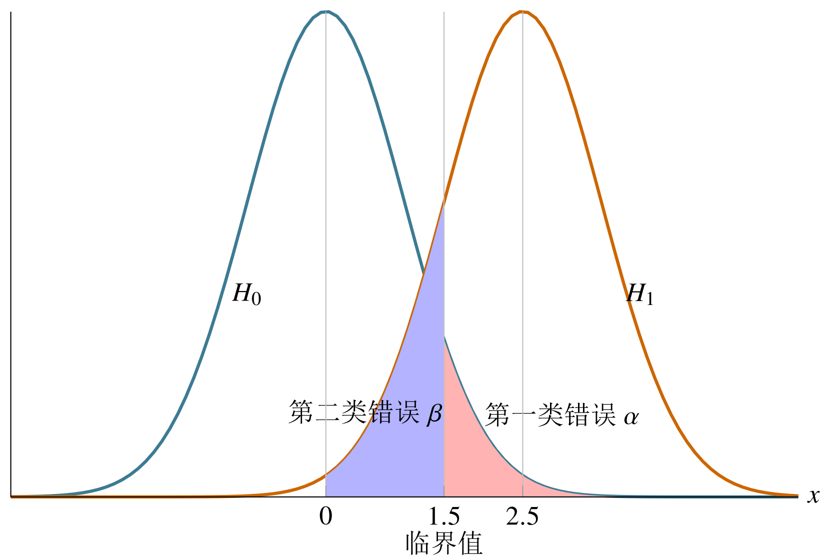
第一类错误：原假设为真时拒绝 $\alpha = \mathrm{P}(\text{拒绝 } \mathrm{H}_0 \mid \mathrm{H}_0)$, 也就是显著性检验。

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

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

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

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

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

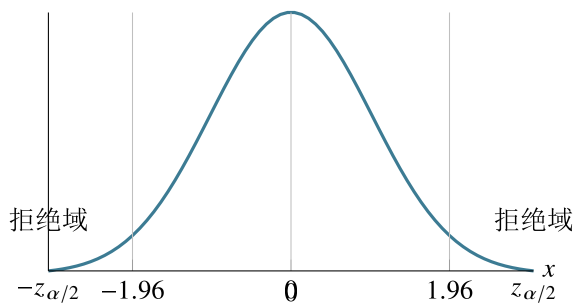
双侧 $2P < \alpha$

单侧 $P < \alpha$

$P < 2P < \alpha$

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”。

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 F 检验

多个检验时可以使用 F 检验

有以下写法：

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

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

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)

证明 如果 H_0 成立，有以下情况

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

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。

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

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

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

$$y=\hat\beta_0+\hat\beta_4x_4+\hat u^r$$

$$y=\hat\beta_0+\hat\beta_1x_1+\hat\beta_2x_2+\hat\beta_3x_3+\hat\beta_4x_4+\hat u^{ur}$$

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

也就是说 $\hat u^r=\hat\beta_1x_1+\hat\beta_2x_2+\hat\beta_3x_3+\hat u^{ur}$

因此可以转换残差平方和 $\sum\hat{u_i^r}^2=\sum(\hat x_i+\hat{u_i}^{ur})^2$

展开这个式子才会用到 $\hat{u_i}^{ur}\hat x_i=0$

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。

```
      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

定理 4.7 (F 检验的等效性)

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

♡

4.8 渐进检验

关于渐进分布4.3。

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。

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：

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 $R_{Q\times P}, Rank(R)=Q$,

$$\sqrt{N}\boldsymbol{R}\left(\widehat{\boldsymbol{\theta}}_N-\boldsymbol{\theta}\right)\xrightarrow{d}Normal\left(0,\boldsymbol{RVR}'\right)$$

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

Delta method：

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

$$\left\{\sqrt{N}\left[\boldsymbol{c}(\widehat{\boldsymbol{\theta}}_N)-\boldsymbol{c}(\boldsymbol{\theta})\right]\right\}'[\boldsymbol{RVR}']^{-1}\{\sqrt{N}\left[\boldsymbol{c}(\widehat{\boldsymbol{\theta}}_N)-\boldsymbol{c}(\boldsymbol{\theta})\right]\} \xrightarrow{d} \chi_Q^2$$

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:

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

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

异方差存在时，只需要修改形式:

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

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

弱外生性：$cov(u,x)=0$, 在此基础上加入 $E(u)=0$, 即可得到 $E(x'x)=0$（这点很容易满足，因为只要有截距项就能满足）。

强外生性：$cov(u,f(x))=0$

具有内生性，也就是至少连弱外生性也没有满足 $cov(u,x)\neq 0$

内生变量 (endogenous variable): 由经济模型内部的其它经济变量所决定的经济变量。

外生变量 (exogenous variable): 由经济模型外部的其它因素所决定的经济变量。

总结: 外生变量是在模型之外决定的, 内生变量是由模型自身决定的。

方法：DID RDD DDD FIX IV Bunching

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 $E(z'u)$

5.2.1 期望

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。

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

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

$$\operatorname{rank} \mathrm{E}(\mathbf{z}'\mathbf{x}) = K$$

rank condition: 秩条件

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

因此得到引理：
这说明多少个内生变量就至少多少个工具变量。也就得到 order condition
order condition:$L \geq K$

$$[\mathrm{E}(\mathbf{z}'\mathbf{x})]\boldsymbol{\beta} = \mathrm{E}(\mathbf{z}'\mathbf{y})$$

同理，

$$\mathrm{E}(\mathbf{x}^{*\prime}\mathbf{x})\boldsymbol{\beta} = \mathrm{E}(\mathbf{x}^{*\prime}y)$$

$$\mathrm{E}(\mathbf{x}^{*\prime}\mathbf{x}) = \Pi'\mathrm{E}(\mathbf{z}'\mathbf{x}) = \mathrm{E}(\mathbf{x}'\mathbf{z})[\mathrm{E}(\mathbf{z}'\mathbf{z})]^{-1}\mathrm{E}(\mathbf{z}'\mathbf{x})$$

$$\Pi = [\mathrm{E}(\mathbf{z}'\mathbf{z})]^{-1}\mathrm{E}(\mathbf{z}'\mathbf{x})$$

$$\mathrm{E}(\mathbf{x}^{*\prime}\mathbf{y}) = \mathrm{E}(\mathbf{x}'\mathbf{z})[\mathrm{E}(\mathbf{z}'\mathbf{z})]^{-1}\mathrm{E}(\mathbf{z}'\mathbf{y})$$

此处 Π 的作用详细见5.4
也就是

$$\boldsymbol{\beta} = [\mathrm{E}(\mathbf{z}'\mathbf{x})]^{-1}\mathrm{E}(\mathbf{z}'y) = [E(zz')^{-1}E(z'X)]^{-1}[E(zz')^{-1}E(z'Y)]$$

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。

$$\hat{\boldsymbol{\beta}} = \left[\left(\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{z}_i\right)\left(\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{x}_i\right)\right]^{-1}\left(\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{z}_i\right)\left(\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{y}_i\right)$$

得到样本矩

$$\hat{\boldsymbol{\beta}}^{IV} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'y_i\right) = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Y}$$

代入 y_i，讨论渐进分布

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{z}_i\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{x}_i\right)\right]^{-1} \cdot \left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{z}_i\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'u_i\right)$$

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

对于**秩条件**的补充：

if $rank(z'x) = k$

prove：$rank[E(x'z)E(x'z)^{-1}E(z'x)] = k$

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。

反证：

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 $E(z'x)v \neq 0$

第二种证明方法，后推前：

已知 $rank(v'E(z'x)E(z'z)^{-1}E(xz')v) = k$, 拆解为 ABA 形式。

$$rank(ABA') = k$$

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

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

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

总结，期望部分对应了

假设 1：$\mathbf{z}'u) = \mathbf{0}$

假设 2：(a) rank E(z′z) = L；(b) rank E(z′x) = K

5.2.2 方差

渐进情况下 $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ 遵循 $N^{-1/2} \sum_{i=1}^{\hat{N}} \mathbf{z}_i' u_i$ 的正态分布。

此时引出**假设 3**：

$$\mathrm{E}(u^2\mathbf{z}'\mathbf{z}) = \sigma^2\mathrm{E}(\mathbf{z}'\mathbf{z}), \text{ where } \sigma^2 = \mathrm{E}(u^2).$$

这个假设等价于 $\mathrm{E}(u^2|\mathbf{z}) = \sigma^2$

因此得到渐进方差：$Avar[\hat{\beta}^{2sls} - \beta] = \sigma^2\{\mathrm{E}(\mathbf{x}'\mathbf{z})[\mathrm{E}(\mathbf{z}'\mathbf{z})]^{-1}\mathrm{E}(\mathbf{z}'\mathbf{x})\}^{-1}$

方差估计为 $\hat{\sigma}^2 \equiv (N - K)^{-1} \sum_{i=1}^{N} \hat{u}_i^2$

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

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 $E(x^* r) = E(\beta z r) = 0$，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：

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

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

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

5.3 2SLS

5.4 工具变量的选择

找工具变量的思路：

相关性 (x) 和排他性 (u)

控制后门：z 对 x 的影响是唯一的，堵住后门。

遥远的历史、空间的平均、自然地理的随机影响

注

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

找工具变量 z

如果 x_2 内生，$E(\hat{\beta}_1) \neq \beta_1, E(\hat{\beta}_1) = \beta_1$

如果 x_2 外生，$E(\hat{\beta}_1) = \beta_1, E(\hat{\beta}_1) = \beta_1$

因此构造工具变量还要用到所有外生变量，因为这个做最有效。

FWL 告诉我们保持其他变量不变的，但一个变量是内生的，会影响其他外生变量的估计。

$$\hat{\beta}_1 = (X_1' M_{x_2} X_1)(X_1' M_{x_2} Y) = \beta_1 + (X_1' M_{x_2} X_1)(X_1' M_{x_2} u)$$

只有当 $x_2' x_1 = 0$ 没有偏相关时，不影响。

偏相关与不相关不同。

好的工具变量，是随机的，影响渠道是唯一的。

先证明 x 对 z 相关，直接回归。

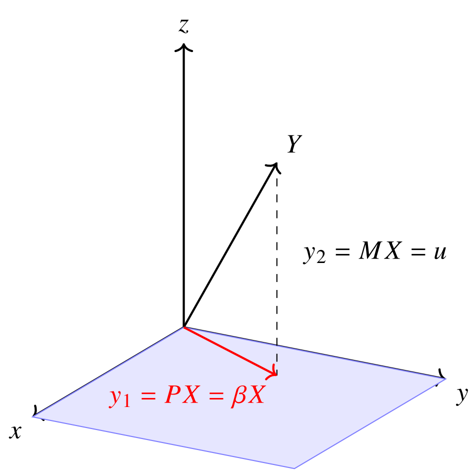
$$\hat{x} = \beta z + u$$

再加入 z, 应该在控制了 x 后，$\beta_2 = 0$

$$y = \beta_0 + \hat{x} + u = \beta_0 + (\beta z + u) + u$$

如果 $L = rank(z'z) > rank(z'x) = k$, 就通过回归降维，找线性映射，此时就有相同的工具变量和解释变量，因此可以求解。也就是 x 对其他外生变量回归。

$x = (x_1, x_2, x_3) \longrightarrow z = (z_1, z_2, x_2, x_3) \longrightarrow (z\hat{\pi}, x_2, x_3)$



x^* 是所有外生变量的线性组合。

当有多个工具变量选择时可以同时使用，因为此时效果更好。

$$\hat{\pi} = z(z'z)^{-1} E(z'x)$$

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

$$\beta_{IV} = \frac{cov\{y,z\}}{cov\{x,z}\}$$

5.5 假设检验

5.5.1 内生性 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

内生性的结果是 $cov(x,u) \neq 0$，因此检验思路是提取基本回归的残差，再加入回归式子进行分析。

1. H0: lnfama – transfer 是外生变量; H1: lnfama – transfer 是内生变量.
2. 进行一阶段回归, 得到残差 vi
3. 将 vi 加入结构方程中进行 OLS 估计.
4. 用 t 检验检验 vi 的系数是否显著不等于 0.

假使 hausman 检验拒绝原假设, 则 lnfama – transfer 是内生变量.

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;
2. Stock & Yogo (2002)——Cragg-Donald chi-squared test

5.5.3 过度识别约束变量

只有一个工具变量, 无法检验该工具变量是否外生。

当工具变量数多于内生解释变量数时, 我们称结构方程存在过度识别。

如果所有的工具变量都满足外生假定, 那么, 结构方程的误差将与所有外生变量 (包括结构方程中的和排除在结构方程外的工具变量) 不相关。

基于上述思想, 可以将 2SLS 估计得到的残差与所有外生

变量做回归来检验工具变量是否外生, 这就是过度识别约束检验检验。

L>K

方程个数大于未知数个数。

因为要检验工具变量的外生性，需要引入多的工具变量做检验。

豪斯曼检验

$H_0 : \beta^{2sls} = \beta^{ols}$

构造 wald 统计量

$$\text{Wald} = \left(\hat{\beta}^{\text{2SLS}} - \hat{\beta}^{\text{OLS}}\right)' \left(\text{Var}(\hat{\beta}^{\text{2SLS}}) + \text{Var}(\hat{\beta}^{\text{OLS}})\right)^{-1} \left(\hat{\beta}^{\text{2SLS}} - \hat{\beta}^{\text{OLS}}\right)$$

豪斯曼检验检验

前提是有效工具变量

对比工具变量和 OLS 检验估计量的差异

问题 5.1 分析下图5.1：

```
                                                      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

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。

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 LATE

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。

一些造成内生性的可能原因：

- 遗漏变量
- 测量误差
  同一时期的变量相互影响，影响系数估计，通常估计偏大。
- 联立相关

思考**遗漏变量**的情况

原回归式子为：

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

估计式为：

$$\hat{\beta}_j = \frac{\Sigma(\hat{x}_{ij} - \bar{x}_j)y_i}{\sum(\hat{x}_{ij} - \bar{x}_j)^2}$$

遗漏变量的情况

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

$$\beta_{no} = E(x_1' x_1)^{-1} E(x_1' y)$$

代入原回归式子即可得到：

$$\beta_{no} = \beta_1 + \beta_2 E(X_1' X_1)^{-1}(X_1 X_2)$$

因此高估和低估取决于两个变量关系，x_1 和 x_2 之间的相关性，β_2 同号高估，异号低估。

思考**联立相关** 的情况

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

$$y_2 = \gamma_0 + \gamma_1 y_1 + v$$

假设 $cov(u_i, v_i) = 0$, 则

$$cov(y_2, u_i) = cov(r_1 y_1 + v_i, u_i) = cov(r_1(\beta_1 y_2 + u_i) + v_i, u_i) = r_1\beta_1 cov(y_2, u_i) + r_1 var(u_i)$$

这里面可以反解出 $cov(y_2, u_i)$

这里使用初级计量的估计式作为参考：

$$\hat{\beta}_1 = \frac{\Sigma(\hat{y}_{i2} - \bar{y}_2)y_i}{\sum(\hat{y}_{i2} - \bar{y}_2)^2} = \frac{\Sigma(\hat{y}_{i2} - \bar{y}_2)u_i}{\sum(\hat{y}_{i2} - \bar{y}_2)^2}$$

关键同样是 $cov(y_2, u_i)$

local average treatment effect 工具变量估计。

导致工具变量估计偏大的一个重要原因是: 工具变量估计的是局部处理效应 (Local Average Treatment Effect, LATE)(工具变量离散情形) 或加权平均局部处理效应 (工具变量连续情形)。

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

$$\beta^{IV} = [E(Z'Z)^{-1}E(Z'X)]^{-1}E(Z'Z)^{-1}E(Z'Y)$$

工具变量增强了一致性，但损失了外部有效性。

第六章 现象面板固定随机效应

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。

时间变动：不是独立同分布。而中心极限要求独立同分布。短面板继续中心极限，长面板则需要时间序列处理的思想。

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

个体效应 c_i 不随时间变动而变动。

6.2 假设

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

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

试图缓解 c_i 的内生性担忧。

6.3 随机效应

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。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。$\sum_i \sum_t x_{it}$

接下来考虑随机效应：

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

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

其中，这里的 s 是指其他时期。
于是得到：

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

$v_i=(v_{i1},v_{i2},\cdots,v_{iT})'$ 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：

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

- **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**：$Rank\ E(X_i'\widehat{\Omega}^{-1}X_i)=k$
- **随机效应假定 3**：(a)$E(u_iu_i'|x_i,c_i)=\sigma_u^2I_T$; (b)$E\left(c_i^2|x_i\right)=\sigma_c^2$

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 $\Omega=\sigma_u^2\cdot I_T+\sigma_c^2\cdot j_Tj_T'$

$$\widehat{\beta}_{RE}=\left(\sum_{i=1}^N X_i'\widehat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^N X_i'\widehat{\Omega}^{-1}y_i\right)$$

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。

$$\Omega^{-1/2}y_i=\Omega^{-1/2}X_i\boldsymbol{\beta}+\Omega^{-1/2}v_i$$

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 $u_i t$, 如果不满足严格外生性，不一定为一致的。

接下来关键就是估计 $\hat{\Omega}$

由于

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

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$\hat{\omega}^2 = \hat{\boldsymbol{\sigma}}_v^2 - \hat{\boldsymbol{\sigma}}_c^2$;
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)

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:

$$y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)\beta + u_{it} - \overline{u}_i$$

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：$E(u_i u_i'|x_i, c_i) = \sigma_u^2 I_T$

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{FE}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}'\ddot{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}'\ddot{y}_{it}\right)$$

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。
接下来考虑渐进分布：
对于

$$y_i = X_i\boldsymbol{\beta} + c_i j_T + u_i$$

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：

$$Q_T y_i = Q_T X_i\boldsymbol{\beta} + Q_T u_i$$

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。

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

$$Var(\hat{\beta}_{FE}) = \sigma_u^2\left(\sum_{i=1}^{N}\ddot{X}_i'\ddot{X}_i\right)^{-1}$$

$$E\left(\ddot{u}_{it}^2\right) = E[(u_{it} - \bar{u}_i)^2] = E\left(u_{it}^2\right) + E\left(\bar{u}_i^2\right) - 2E(u_{it}\bar{u}_i)$$
$$= \sigma_u^2 + \frac{\sigma_u^2}{T} - 2\frac{\sigma_u^2}{T} = \sigma_u^2\left(1 - \frac{1}{T}\right)$$

对 t 求和

$$\sum_{t=1}^{T}E\left(\ddot{u}_{it}^2\right) = \sigma_u^2(T-1)$$

对 n 求和

$$[N(T-1)]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(\ddot{u}_{it}^2\right) = \sigma_u^2 E\left([N(T-1)]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{u}_{it}^2\right) = \sigma_u^2$$

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)-k}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2$$

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。
异方差时，估计量为：

$$\widehat{Var(\beta_{FE})} = \left(\sum_{i=1}^{N}\ddot{X}_i'\ddot{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\ddot{X}_i'\hat{\ddot{u}}_i\hat{\ddot{u}}_i'\ddot{X}_i\right)\left(\sum_{i=1}^{N}\ddot{X}_i'\ddot{X}_i\right)^{-1}$$

注 第二种方法——把 c_i 看作参数进行估计，使用截距估计分解。满足三个假设时，估计相同。可以利用 FWL 进行证明。（Green 7 版 360-361）

一阶差分

$$y_{it}-y_{it-1}=(x_{it}-x_{it-1})\boldsymbol{\beta}+u_{it}-u_{it-1}$$

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

假定 $1E\left(u_{it}|x_i,c_i\right)=0$ 假定 $2Rank(\sum_{t=2}^{T}E(\Delta x_{it}'\Delta x_{it}))=k$ 假定 $3E(\Delta u_i'\Delta u_i|x_i,c_i)=\sigma_{\Delta u}^2 I_{T-1}$

$$\hat{\beta}_{FD}=\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}'\Delta x_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}'\Delta y_{it}\right)$$

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，

$$Var\widehat{\left(\hat{\beta}_{FD}\right)}=\left(\sum_{i=1}^{N}\Delta X_i'\Delta X_i\right)^{-1}\left(\sum_{i=1}^{N}\Delta X_i'\widehat{\Delta u}_i\widehat{\Delta u}_i'\Delta X_i\right)\left(\sum_{i=1}^{N}\Delta X_i'\Delta X_i\right)^{-1}$$

向前正交变换 stata 命令 xtabond2

6.5 比较

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。

Pooled OLS: $E(x_{it}'v_{it})=0$, 同期外生。

Random Effect: $E(x_{is}'v_{it})=0,\forall s,t$, 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

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。

注 两期面板：FE=FD（考过）

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{FE}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}'\ddot{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}'\ddot{y}_{it}\right)$$

$$\hat{\beta}_{FD} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}'\Delta x_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}'\Delta y_{it}\right)$$

$$\sum_{t=1}^{2}\ddot{x}_{it}'\ddot{x}_{it} = \sum_{t=1}^{2}\ddot{x}_{it}^2 = \sum_{t=1}^{2}(x_{it}-\bar{x}_i)^2$$

在两期中，组间均值为 $\bar{x}_i = \frac{x_{i1}+x_{i2}}{2}$

$$\begin{aligned}\sum_{t=1}^{2}(x_{it}-\bar{x}_i)^2 &= \sum_{t=1}^{2}(x_{it}-\frac{x_{i1}+x_{i2}}{2})\\ &= \frac{(x_{i2}-x_{i1})^2}{2}\end{aligned}$$

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。

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。

$$(\hat{\beta}_2 - \hat{\beta}_1)'[var(\hat{\beta}_2) - var(\hat{\beta}_1)]^{-1}(\hat{\beta}_2 - \hat{\beta}_1)$$

服从 χ 分布

$var((\hat{\beta}_2 - \hat{\beta}_1))$ 依赖于原始方程误差 u_{it} 的协方差。序列无关时，协方差为 0。

豪斯曼检验选择 re 或者 fe 的方差之一作为检验。

单个检验所转化为 t 检验。

$$t = (\hat{\beta}_{FE} - \hat{\beta}_{RE}) / \left\{ \left[se(\hat{\beta}_{FE}) \right]^2 - \left[se(\hat{\beta}_{RE}) \right]^2 \right\}^{\frac{1}{2}}$$

6.6 检验

6.6.1 思路 1

$$\text{随机效应存在} \rightarrow E(v_{is}v_{it}) = \sigma_c^2, \forall s \neq t$$

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。

$$\sqrt{N}\widehat{\sigma}_c^2 \sim Normal\left(0, E\left[\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\tilde{v}_{is}\tilde{v}_{it}\right)^2\right]\right)$$

Listing 6.1: 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。

6.6.2 思路 2

从如下条件出发：

$$H_0{:}E\left(c_i|x_i\right) = 0; H_1{:}E\left(c_i|x_i\right) \neq 0$$

构造 wald 统计量，

$$H=(\hat{\beta}_{FE}-\hat{\beta}_{RE})'\left[\widehat{Var(\hat{\beta}_{FE})}-\widehat{Var(\hat{\beta}_{RE})}\right]^{-1}(\hat{\beta}_{FE}-\hat{\beta}_{RE})$$

也可以选择 t 统计量

$$t=(\hat{\beta}_{FE}-\hat{\beta}_{RE})/\left\{\left[se(\hat{\beta}_{FE})\right]^2-\left[se(\hat{\beta}_{RE})\right]^2\right\}^{\frac{1}{2}}$$

6.6.3 思路 3

基于回归的检验。思路: 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。因此, 可以构造如下回归方程来检验:

$$\check{y}_i=\check{X}_i\boldsymbol{\beta}+\ddot{w}_i\xi+\check{v}_i$$

$\check{h}_i=\widehat{\Omega}^{-\frac{1}{2}}h_i$,$\ddot{W}_i$ 是 $\ddot{x}_i$(所有组内去均值的解释变量) 的一个子集。

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。

Listing 6.2: 检验

```
xtreg y x, fe est store fe

xtreg y x, re est store re

hausman fe re, sigmamore/sigmaless
```

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:

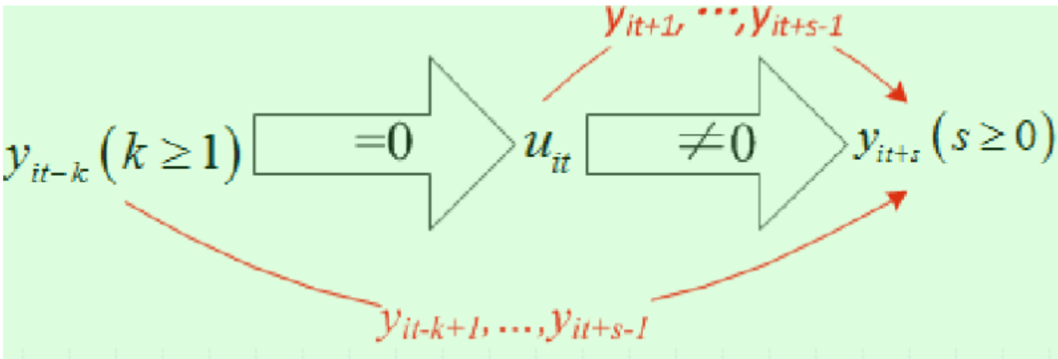


图 7.1: 动态完备

$$\operatorname{var}(\hat{\beta}_{FEIV}) = \left[\left(\sum_{i=1}^{N}\ddot{X}_i'\ddot{Z}_i\right)\left(\sum_{i=1}^{N}\ddot{Z}_i'u_iu_i'\ddot{Z}_i\right)^{-1}\left(\sum_{i=1}^{N}\ddot{Z}_i'\ddot{X}_i\right)\right]^{-1}$$

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：

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

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。♣

$$E(y_{it} \mid x_{it}, x_{it-1}, \ldots, x_{i1}, c_i) = E(y_{it} \mid x_{it}, c_i) = x_{it}\beta + c_i$$

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：

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

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：

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

不考虑样本容量:

$$\arg\min_{\widehat{\beta}}\left[Z'(y-X\hat{\beta})\right]'\widehat{W}\left[Z'(y-X\hat{\beta})\right]$$

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。

$$
\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{p}(CWC')^{-1}CW\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(z_i'u_i)\right]+o_p(1)
$$

$$
\operatorname{Avar}(\hat{\beta})=\frac{1}{N}(CWC')^{-1}CW\cdot\Lambda\cdot WC'(CWC')^{-1}
$$

$$
\Lambda\equiv E(z_i'u_iu_i'z_i)=\operatorname{var}(z_i'u_i)=\frac{1}{N}\sum_{i=1}^{N}z_i'u_iu_i'z_i
$$

u_i 是向量。

8.2 GMM 与 2SLS 的关系

有权重矩阵 $\widehat{W}=\left(N^{-1}\sum_{i=1}^{N}z_i'z_i\right)^{-1}=(Z'Z/N)^{-1}$

$$
\begin{aligned}
\hat{\beta}_{GMM} &= (X'Z(Z'Z/N)^{-1}Z'X)^{-1}(X'Z(Z'Z/N)^{-1}Z'y) \\
&= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'y) \\
&= (X^{*\prime}X)^{-1}(X^{*\prime}y)=\hat{\beta}_{2SLS}
\end{aligned}
$$

$$
\operatorname{Avar}(\hat{\beta})=\frac{1}{N}(C\Lambda^{-1}C')^{-1}
$$

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：

$$
\tilde{\beta}=\hat{\beta}_{\text{2SLS}}\to\widehat{W}=(Z'Z/N)^{-1}
$$

$$\hat{\Lambda} = N^{-1} \sum_{i=1}^{N} z_i' \tilde{u}_i \tilde{u}_i z_i$$

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**。

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 $Y_i^0 = \mu_i^0 + v_i^0$，　$Y_i^1 = \mu_i^1 + v_i^1$，

即可得到：

$$Y_i = \mu_i^0 + \left(\mu_i^1 - \mu_i^0\right) D_i + D_i\left(v_i^1 - v_i^0\right) + v_i^0$$

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 ATT 来说，

$$ATT = E\left(\mu_i^1 - \mu_i^0 | D_i = 1\right) + E\left(v_i^1 - v_i^0 | D_i = 1\right)$$

$$ATT = E(Y_i|D_i = 1) - E(Y_i|D_i = 0) - \left[E(v_i^0|D_i = 1\right) - E\left(v_i^0|D_i = 0)\right]$$

代入相关的条件期望式子为：

$$\begin{aligned}
&E(Y_i|D_i = 1) - E(Y_i|D_i = 0) \\
&= ATT + E\left(v_i^0|D_i = 1\right) - E\left(v_i^0|D_i = 0\right)
\end{aligned}$$

因此其无偏估计的条件假设为 $E(v_i^0|D_i) = 0$。

假定的含义: 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, 没有**选择偏误** (Selection bias) 。

反事实：$Y_i^0|D_i = 1, Y_i^1|D_i = 0$

对于 ATE 来说，

无偏估计需要满足：

$$E\left(u_i|D_i\right) = 0 \Rightarrow \boldsymbol{A}.E\left(v_i^0|D_i\right) = 0; \boldsymbol{B}.E\left[v_i^1 - v_i^0|D_i\right] = 0$$

如果式子还有控制变量，则需要假定：

$$Y_i^1, Y_i^0 \perp D_i \mid X_i$$

假定含义:

(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。

在调价按假定下，

$$\begin{aligned}
ATT &= E\left(Y_i^1 - Y_i^0 \mid D_i = 1, X\right) \\
&= E\left(Y_i|D_i = 1, X\right) - E\left(Y_i|D_i = 0, X\right) \\
&= E\left(Y_i^1 \mid X\right) - E\left(Y_i^0 \mid X\right) = ATE
\end{aligned}$$

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。

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:

$$Y_{it-1} = Y_{it-1}^0 = \mu_i^0 + c_i + \rho\varphi_{t-1} + \varepsilon_{it-1}^0$$

$$Y_{it} = \mu_i^0 + \left(\mu_i^1 - \mu_i^0\right) D_{it} + \left(\varepsilon_{it}^1 - \varepsilon_{it}^0\right) D_{it} + c_i + \rho\varphi_t + \varepsilon_{it}^0 = \alpha + \beta^{ATT} D_{it} + c_i + \rho\varphi_t + u_{it}$$

$$\Delta Y_{it} = \beta^{ATT} \Delta D_{it} + \rho\Delta\varphi_t + \Delta u_{it}$$

$$\Delta D_{it} = D_i \times 1 - D_i \times 0 = D_i$$

$$\begin{aligned}
\beta^{\mathrm{ATT}} &= E(\Delta Y_{it} \mid D_i = 1) - E(\Delta Y_{it} \mid D_i = 0) \\
&= E(Y_{it} \mid D_i = 1) - E(Y_{it-1} \mid D_i = 1) - [E(Y_{it} \mid D_i = 0) - E(Y_{it-1} \mid D_i = 0)] \\
&= \underbrace{E(Y_{it}^1 \mid D_i = 1) - E(Y_{it-1}^0 \mid D_i = 1)}_{\text{first difference}} - \underbrace{\left[E(Y_{it}^0 \mid D_i = 0) - E(Y_{it-1}^0 \mid D_i = 0)\right]}_{\text{second difference}}
\end{aligned}$$

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。

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：

$$Y_{it} = \alpha + \beta^{ATT} D_{it} + X_{it}\gamma + c_i + \rho\varphi_t + u_{it}$$

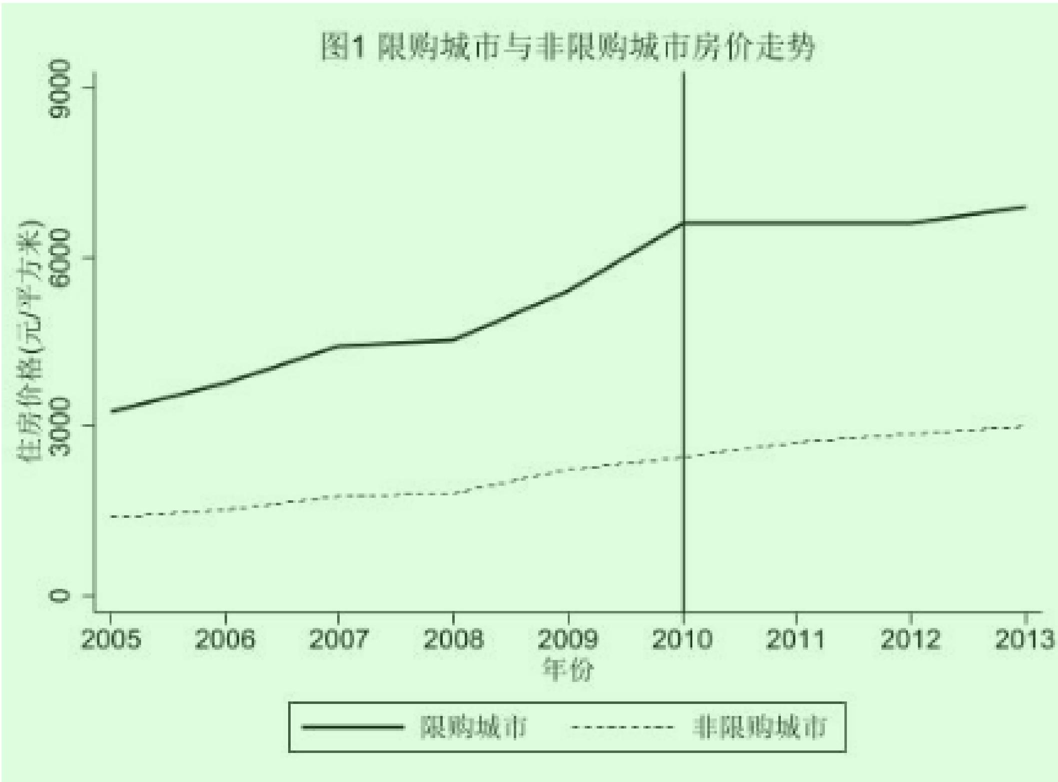


图 **9.1:** 上偏不满足平行趋势检验的例子

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。

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。

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.

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)

ln(hp)	Coef.	Robust 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 9.2: 回归结果

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%.

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 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。