Econometrics

数据形态

截面数据 Cross-Section

- 各个观察值间多多少少是独立的
- 截面不一定是随机抽样的，比如受访者拒绝响应、样本聚类
- 典型用于应用微观研究

时间序列 Time Series

- 几乎可以肯定有序列相关
- 典型特征是趋势和季节性
- 典型用于应用宏观和金融研究

池化截面 Pooled Cross-Sections

不同于面板数据，是两个时间点做两次抽样，**不要求两次抽样的样本是同一个群体**

- 经常用于评估政策效果（DID）

面板数据 Panel/Longitude

不同于聚合截面，两个时间点的记录样本必须是同一个群体

- 经常用于对滞后响应建模

网络数据 Network

可能存储于图式数据库，结构化后可能会类似如下形式

Year	Export Country Code	Import Country Code	Product Code	Value	QTY
2016	4	12	80132	26.313	3
2017	12	4	130190	11.507	5

样本和总体的概念

即便拿到了全样本，也不能说是拿到了总体（population）。因为全样本也只是一个“实现”，还有“未实现的可能性”

因果推断

解决了内生性问题就可以认为是“因果”关系？

数学回顾

1. 形如 $\mathbf{x}^T[Var(\mathbf{x})]^{-1}\mathbf{x}$ 的二次型：直观含义是以 $[Var(\mathbf{x})]^{-1}$ 为权重，将 $\mathbf{x}$ 到零向量 $\mathbf{0}$ (原点)的距离标准化。一维情形下，化为 $(\frac{x}{\sqrt{Var(x)}})^2$，即 x 离原点有多少个标准差距离（马氏距离）的平方。
2. 夹心估计量，$Var(AX) = AVar(X)A^T$。其中 A 为对称矩阵
3. $Var(X) = E(XX^T) - E(X)[E(X)]^T$
4. 正态分布峰度为3，峰度越大越厚尾
5. 期望迭代

$$
\begin{aligned}
E_X[E(Y|X)] &= \int_{-\infty}^{+\infty} f(x)E(Y|X)dx && (1)\\
&= \int_{-\infty}^{+\infty} f(x)\int_{-\infty}^{+\infty} y\frac{f(x,y)}{f(x)}dydx && (2)\\
&= \int_{-\infty}^{+\infty} ydy\int_{-\infty}^{+\infty} f(x,y)dx && (3)\\
&= \int_{-\infty}^{+\infty} yf(y)dy && (4)\\
&= EY && (5)
\end{aligned}
$$

6. 相互独立⇒均值独立⇒线性不相关。且均值独立不是对两个变量对称的关系
7. 多维正态分布
   1. 二维正态

$$
f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}exp\{-\frac{1}{2(1-\rho^2)}[(\frac{x_1-\mu_1}{\sigma_1})^2 - \frac{2(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + (\frac{x_2-\mu_2}{\sigma_2})^2]\} \quad (6)
$$

   2. 多维矩阵表达

$$f(x_1,\cdots,x_n)=\frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}}exp\{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)\} \tag{7}$$

$$\text{其中，}\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_2\sigma_1 & \sigma_2^2\end{pmatrix}$$

$$|\Sigma|=(1-\rho^2)\sigma_1^2\sigma_2^2$$

$$\Sigma^{-1}=\begin{pmatrix}\frac{1}{(1-\rho^2)\sigma_1^2} & -\frac{\rho}{(1-\rho^2)\sigma_1\sigma_2}\\ -\frac{\rho}{(1-\rho^2)\sigma_2\sigma_1} & \frac{1}{(1-\rho^2)\sigma_2^2}\end{pmatrix}$$

小样本OLS

所谓“小样本”其实更应该叫做任意样本量，反正就是没有 $lim\ n\to\infty$

整体模型，线性假定 $\mathbf{y}=X\beta+\epsilon$，注意矩阵 X 的第一列元素 $x_{i1}=1$，原因是常数项

最小化残差平方和，注意 ϵ 为误差，$\mathbf{e}$ 为残差

$$\begin{aligned}\min_{\tilde\beta} SSR(\tilde\beta)&=\sum_{i=1}^n e_i^2=\mathbf{e}^T\mathbf{e} &(8)\\ &=(\mathbf{y}-X\tilde\beta)^T(\mathbf{y}-X\tilde\beta) &(9)\\ &=\mathbf{y}^T\mathbf{y}-\mathbf{y}^TX\tilde\beta-\tilde\beta X^T\mathbf{y}+\tilde\beta^TX^TX\tilde\beta &(10)\\ &=\mathbf{y}^T\mathbf{y}-2\mathbf{y}^TX\tilde\beta+\tilde\beta^TX^TX\tilde\beta &(11)\end{aligned}$$

其中，$\mathbf{y}$ 为 $n*1$ 向量，X 为 $n*K$ 矩阵，β 为 $K*1$ 向量 可以看到各项均为内积形式的 $1*1$ 常数

一些基本矩阵微分规则

这里都是按列向量求导的梯度微分 ∇

$$\text{内积微分，}\frac{\partial(\alpha^T\beta)}{\partial\beta}=\alpha \tag{12}$$

$$\text{二次型微分，}\frac{\partial(\beta^TA\beta)}{\partial\beta}=2A\beta\text{，其中}A\text{为对称矩阵}$$

使用上面的微分规则，对残差平方和目标函数求一阶导

$$\frac{\partial(SSR)}{\partial\tilde\beta}=-2X^T\mathbf{y}+2X^TX\tilde\beta=0 \tag{13}$$

从而有正规方程组：

$$(X^TX)_{K*K}\tilde\beta_{K*1}=X^T_{K*n}\mathbf{y}_{n*1} \tag{14}$$

$$
\begin{aligned}
&\text{移项有} X^T\mathbf{y}-(X^TX)\tilde{\beta}=\mathbf{0}\\
&\text{提公因式，} X^T(\mathbf{y}-X\tilde{\beta})=\mathbf{0}\\
&\text{亦即} X^T\mathbf{e}=\mathbf{0}\text{，残差和解释变量正交}
\end{aligned}
\tag{15}
$$

因 $x_{i1}=1$, 转置后 X^T 第一行元素全为 1，从而上式也意味着 $\sum_{i=1}^n e_i=0$

最终得到

$$
\begin{aligned}
&\tilde{\beta}=(X^TX)^{-1}X^T\mathbf{y}\\
&\hat{\mathbf{y}}=X\tilde{\beta}=X(X^TX)^{-1}X^T\mathbf{y}=\mathbf{P}\mathbf{y}
\end{aligned}
\tag{16}
$$

其中 $\mathbf{P}=X(X^TX)^{-1}X^T$ 称为投影矩阵，用 $\mathbf{P}$ 左乘任何向量就得该向量在超平面 X 上的投影 另一方面 $\mathbf{e}=\mathbf{y}-\hat{\mathbf{y}}=\mathbf{y}-\mathbf{P}\mathbf{y}=(\mathbf{I}_n-\mathbf{P})\mathbf{y}=\mathbf{M}\mathbf{y}$ 其中 $\mathbf{M}=\mathbf{I}_n-\mathbf{P}$ 为消灭矩阵，用它左乘任何向量就是该向量对超平面 X 投影后的残差向量

平方和分解公式

离差平方和 Total Sum of Squares (TSS)

$$
\sum_{i=1}^n (y_i-\bar{y})^2 \tag{17}
$$

Explained Sum of Squares (ESS)

$$
\sum_{i=1}^n (\hat{y}-\bar{y})^2 \tag{18}
$$

残差平方和 Residual Sum of Squares (RSS)

$$
\sum_{i=1}^n e_i^2 \tag{19}
$$

分解有

$$
TSS=ESS+RSS \tag{20}
$$

拟合优度

$$
R^2=\frac{ESS}{TSS} \tag{21}
$$

校正拟合优度

$$
\bar{R^2}=1-\frac{\sum_{i=1}^n e_i^2/(n-K)}{\sum_{i=1}^n (y_i-\bar{y})^2/(n-1)} \tag{22}
$$

通过自由度来惩罚模型复杂度

一个高的 R^2 对于模型预测精度有利，但不能证明某些变量的因果关系很强。所以现代研究一般只是按惯例列报，不怎么强调 R^2 的具体讨论了

古典线性回归模型关键假定

1. 线性假定，真实DGP线性于参数
2. 严外生性 $E(\varepsilon_i|X)=0$，即扰动项和所有解释变量均不相关，$Cov(\varepsilon_i, x_{jk})=0, \ \forall j,k$。用期望迭代即可有 $E(\varepsilon_i)=0$
3. 不存在严格多重共线性，亦即 X 满列秩。进而 X^TX 正定、可逆，因为 $r(X^TX)=r(X)$，对称且满秩（这里 X 不一定是方阵，所以谈不上 X 可逆，实际上不能直接用分解证明其正定）。此性质是回归能够算得出来的必要条件
4. 球形扰动：同方差、无自相关，亦即 $Var(\varepsilon|X)=\sigma^2\mathbf{I}_n$，这里右边的表达式中不含 X，也就隐含了 ε 二阶矩独立于 X
5. 扰动正态：$\varepsilon|X\sim N(\mathbf{0},\sigma^2\mathbf{I}_n)$，同样这里右边的分布表达式中不含 X，所以隐含了 ε 和 X 是独立的。这个假定也就包含了球状扰动

可以这样理解这些古典假定：手上有了 $\mathbf{y}$ 和 X，希望把它们投到超平面上 $\hat{\mathbf{y}}=X\hat{\beta}+\mathbf{e}$

首先要让这个超平面存在，不只是像 SVM 那样插入一块作为分隔。所以要有线性假定，真实 DGP 线性于参数，然后希望有如下的合意性质。

合意性质

1. 线性性，$\hat{\beta}$ 可以视为 $\mathbf{y}$ 的线性组合。由线性假定所得
2. 无偏性。由严外生性所得
3. 协方差矩阵形式简洁，$Var(\hat{\beta}|X)=(X^TX)^{-1}X^TVar(\varepsilon|X)X(X^TX)^{-1}=\sigma^2(X^TX)^{-1}$。由球形扰动所得
4. BLUE中的方差最小。同样有球形扰动所得
5. 假设检验和统计推断。由扰动正态支持

特别地，对比伍德里奇的经典线性模型假定（CLM） MLR.1~6。前五个也叫高斯马尔科夫假定

> MLR.1 线性于参数
>
> MLR.2 随机抽样（陈强的版本应该是隐含了）
>
> MLR.3 不存在完全共线性
>
> MLR.4 零条件均值 $E(u|X)=0$（就是严外生性）
>
> MLR.5 同方差性 $Var(u|X)=\sigma^2$ （就是球状扰动）

> MLR.6 误差 u 独立于解释变量 X，并且 $u \sim N(0, \sigma^2)$（这里其实也和陈强的版本一致，要求 u 和 X 独立）

假设检验/分类标准的术语

			标准集	
		正样本	负样本	合计
	正样本	TP	FP	P
测试集	负样本	FN	TN	N
	合计	T	F	$P+N=T+F$

1. Precision 查准率 = TP/P
2. Recall 召回率（真阳性率/敏感性） = TP/T
3. 假阳性率/特异性 = FP/F
4. Accuracy 准确率 = $(TP+TN)/(T+F)$
5. F 分数 = $\frac{2}{\frac{1}{Reacll}+\frac{1}{Precision}}$，即查准率和召回率的调和平均数
6. ROC Curve：假阳性率为横轴，真阳性率为纵轴
7. PR Curve：Recall 为横轴，Precision 为纵轴
8. 显著性水平 = FN/T = α
9. Power 检验的势 = TN/F = $1-\frac{FP}{F}$ = $1-\beta$

对数/半对数的典型系数解释

$$\widehat{log(wage)} = \beta_0 + \beta_1 educ + u \tag{23}$$

其中 $\beta_1 = \frac{\partial \widehat{log(wage)}}{\partial educ} = \frac{\frac{\partial wage}{wage}}{\partial educ}$，分子部分是百分比。所以解释是 educ 每增加一个单位，wage 增加 $100\beta_1\%$。**注意半对数一方百分号内要乘100**

$$\widehat{log(wage)} = 0.584 + 0.083educ \tag{24}$$

这个具体例子中，educ 每增加一单位，wage 增加8.3%

$$\widehat{log(salary)} = 4.822 + 0.257log(sales) \tag{25}$$

都是对数的情况下，sales 增加 1%，salary 增加 0.257%。**百分号内不用再乘 100**

一般来说金额、人口都会取对数，其他一些比较大的数字也会取。名义上是为了识别非线性关系、弹性关系，一定程度上消除量纲影响。实际操作上，**取对数更容易显著**。

大样本OLS

小样本需要严外生和正态扰动的假设太强；即便有假设成立都还需要推导精确分布。所以放松要求，换用大样本（$n \geq 30$)

这里的“大样本”混合了一部分时间序列的内容

随机收敛

依均方收敛 $\xrightarrow{chebyshov\text{不等式}}$ 依概率收敛 $\Rightarrow$ 依分布收敛

Khinchin 大数定律，由依均方收敛推出的依概率收敛

Levy-Lindberg 中心极限定理，依分布收敛

随机过程

严平稳：要求相对距离下有限维分布不变，而不取决于实际的绝对距离，自然也就有单个同分布。但不对跨期是否存在序列相关做要求。此性质主要用来代替同分布的要求

弱平稳：相比严平稳不要求同分布，只要求到二阶矩（期望、方差、协方差）相同即可

渐进独立性/遍历性/弱相依/各态历经性：允许存在序列相关，只要这种相关关系在极限处消失即可。只要相距够远，就可以视为近似独立（严格来说是不相关）。主要用来代替独立的要求

渐近独立定理

渐近独立的严平稳过程的样本均值是总体均值的一致估计。

配合连续函数变换，也就推广了矩法估计。**只要是渐近独立严平稳，样本矩都是总体矩的一致估计**

仅要求：

- 严平稳
- 渐近独立

也就是说解除了独立的要求，但还是要求同分布

推广了：

- 大数定律
- 中心极限定理

- 矩估计法

大样本假定

1. 线性假定，和小样本一样
2. (K+1)维随机过程 $y_i, x_{i1}, \cdots, x_{iK}$ 为渐近独立严平稳过程，从而适用大数定律和中心极限定理
3. 前定解释变量，同期外生。$E(x_{ik}\varepsilon_i)=0, \forall i, k$ ，所谓“前定”说的是仿佛在扰动项产生前解释变量就已产生，只是要求同期外生，不是说跨期也要不相关。放松了严外生
4. 秩条件，不存在严格多重共线性。和小样本一样
5. 不再要求球形扰动，大样本下渐近方差支持异方差稳健（自相关要另用其他办法）
6. 不再要求扰动正态，因为已有渐近正态性质

合意性质

1. 一致性，由前定解释变量（同期外生）所得
2. 渐近正态性，由渐近独立严平稳所得
3. 统计推断使用标准正态分布代替 t 分布，χ^2 分布代替 F 分布，由渐进正态推得

特别地，对比伍德里奇的时间序列假设 TS.1~6。因为伍德里奇的时间序列一开始没有大样本渐进推断，所以假设要求比较严格

> TS.1 线性于参数
>
> TS.2 无完全共线性
>
> TS.3 零条件均值： $E(u_t|X)=0$
>
> TS.4 同方差： $Var(u_t|X)=Var(u_t)=\sigma^2$
>
> TS.5 无序列相关：给定 X，任意两期的误差都不相关。$Cov(u_t, u_s|X)=0, \forall t \neq s$
>
> TS.6 正态性：误差 u_t 独立于解释变量 X，并且 $u \sim N(0, \sigma^2)$

考虑上大样本的渐近版本 TS.1'~6'

> TS.1' 线性和弱相关：要求随机过程 $\{(x_{t1}, x_{t2}, \cdots, x_{tk}, y_t : t=1,2,\cdots,n\}$ 是平稳的、弱相关的。而且服从线性模型。（也就是陈强版本的渐近独立严平稳）
>
> TS.2' 无完全共线性
>
> TS.3' 同期外生：$E(u_t|x_{t1}, x_{t2}, \cdots, x_{tk})=0$，放松了 TS.3
>
> TS.4' 同期同方差：$Var(u_t|\mathbf{x}_t)=\sigma^2$，陈强版本因为大样本支持渐近方差也放松了这个
>
> TS.5' 无序列相关：$E(u_t u_s|\mathbf{x}_t, \mathbf{x}_s)=0, \forall t \neq s$，陈强版本似乎只要渐近独立（？）

时间序列典例

AR（1）

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{26}$$

其中 $\{\varepsilon_t\}$ 为 i.i.d，且 $cov(y_{t-1}, \varepsilon_t) = 0$

- 当 $\rho = 1$ 时，$\{y_t\}$ 不平稳，随机游走
- 当 $\rho < 1$ 时，$\{y_t\}$ 严平稳

白噪音

对于弱平稳过程 $\{x_t\}$，如果对 $\forall t$ 都有 $E(x_t) = 0$ 且 $cov(x_t, x_{t+k}) = 0, (\forall k \neq 0)$ 则为白噪音。直观即每一期期望都保持为零，不同期序列不相关，且各期直到二阶矩都保持相同（当然也就有方差不变 $var(x_t) = \sigma^2$）

异方差

1. 球形扰动的主对角线元素不完全相等
2. 后果
    1. OLS仍然无偏、一致且渐进正态
    2. 估计量方差不再是同方差假设下的，t检验、F检验失效
    3. 高斯马尔科夫定理不再成立
3. 检验手段
    1. BP 检验
    2. White 检验
4. 处理方法
    1. OLS + 稳健标准误
    2. 加权最小二乘法（WLS）
    3. 可行加权最小二乘法（FWLS）

自相关

1. 球形扰动的非主对角线元素不全为0
2. 后果
    1. OLS仍然无偏、一致且渐进正态
    2. 估计量方差不再是同方差假设下的，t检验、F检验失效
    3. 高斯马尔科夫定理不再成立
3. 检验手段
    1. BG检验
    2. Q检验
    3. DW检验

4. 处理方法
    1. OLS+异方差自相关稳健标准误
    2. 准差分法
    3. 广义最小二乘法（GLS）
    4. 修改模型设定

模型设定和数据问题

遗漏变量

$$\begin{aligned} wage &= \beta_0 + \beta_1 educ + \beta_2 abli + u \\ abil &= \delta_0 + \delta_1 educ + v \\ wage &= (\beta_0 + \beta_2\delta_0) + (\beta_1 + \beta_2\delta_1) educ + (\beta_2 v + u) \end{aligned} \tag{27}$$

1. 后果
    1. 如果被遗漏变量和解释变量不相关，那么扰动项还是和解释变量不相关。从而 OLS 仍然一致，但**可能**因为扰动项中多了一项，从而其方差 σ^2 增大了，进而系数方差有所增大。可能会导致系数不显著
    2. 如果被遗漏变量和解释变量相关，那么扰动项也和解释变量相关。大样本下 OLS 不一致。高估还是低估的方向在两个变量的情况下比较容易推测，但变量数量一多就难了
        1. 系数高估：(27) 式中 $\beta_2\delta_1$ 为正，亦即两者同号
        2. 系数低估：(27) 式中 $\beta_2\delta_1$ 为负，亦即两者异号
2. 处理方法
    1. 加入尽可能多的控制变量。但也别乱加，见下面加入无关变量的问题
    2. 随机试验和自然实验
    3. 工具变量法
    4. 使用面板数据

加入无关变量

1. 后果：OLS 仍然一致，但是解释变量系数 $\hat{\beta}$ 的方差一般会增大，从而可能不显著
2. 处理方法：别乱加变量，小心加剧共线性问题，最好遵循理论指导。但这里就有建模策略是从小到大还是从大到小的问题，结合ML有lasso选择变量方法

解释变量个数的选择

1. 奥卡姆剃刀原则
2. 可用的权衡准则
    1. 校正可决系数 $\bar{R}^2$
    2. 赤池（Akaike）信息准则
    3. 贝叶斯信息准则
    4. 从大到小的序贯t规则

对函数形式（模型设定）的检验

1. Ramsey的RESET检验

（非严格）多重共线性

1. 后果：
    1. 高斯马尔科夫定理仍然成立，OLS仍然 BLUE。但这个 BLUE 只是说线性无偏估计量里方差最小，可能其他估计量可以更小
    2. 矩阵(X^TX)变得“几乎”不可逆，$(X^TX)^{-1}$变得很”大”，使得方差 $Var(\hat{\beta}|X) = \sigma^2(X^TX)^{-1}$ 增大，系数估计变得不准确。 X 中的元素轻微的变化都会引起很大的变化
    3. 典型症状是整个方程的 R^2 较大， F 检验也显著，但是单个系数的 t 检验却不显著。而且增减解释变量会使原有系数估计值产生较大变化（增加的变量和原来的变量多重共线性）。直观上就是无法区分高度相关的解释变量对被解释变量的单独影响力
    4. 具体来说，某一个具体的 $\hat{\beta}_k$ 的方差公式为 $Var(\hat{\beta}_k|X) = \dfrac{\sigma^2}{(1-R_k^2)S_k}$。其中
        1. R_k^2 为第 k 个解释变量 x_k 对其他所有剩余解释变量回归的 R^2，如果存在多重共线性，这个 R_k^2 会非常接近 1
        2. $S_k \equiv \sum_{i=1}^n (x_{ik} - \overline{x}_k)^2$ 为 x_k 自己的离差平方和
2. 检验手段：
    1. R_k^2与方差膨胀因子 $VIF \equiv \dfrac{1}{1-R_k^2}$
3. 处理方法：
    1. 如果不关心具体的回归系数，只关心整个方程的预测能力（ML倾向），那大可不必理会多重共线性
    2. 如果重点讨论的变量已经显著，同样可以不理会。因为方差膨胀下都显著，那么不膨胀（没有多重共线性）就会更显著
    3. 如果重点讨论变量不显著，就要尝试处理：
        1. 加大样本容量
        2. 剔除多重共线性变量

    3. 将变量标准化，减均值除标准差

极端数据离群值

1. 检验手段：遍历去除第 i 个观测值，比较回归系数。计算杠杆作用 lev
2. 处理方法：截尾、缩尾

虚拟变量

1. 注意虚拟变量陷阱：不能放入所有类别，否则将出现严格多重共线性

结构变动检验

1. 结构变动日期已知：邹检验。全样本、前后总共分三次回归
2. 结构变动日期未知：匡特似然比（QLR）

缺失数据与线性插值

变量单位选择

1. 变量间数量级不要过于悬殊，影响到 $(X^TX)^{-1}$ 的计算

工具变量法

1. 内生性：
    1. 解释变量和扰动项相关。违背前定变量或者叫同期外生的假设
    2. 来源可以是
        1. 遗漏变量
        2. 联立方程双向因果
        3. 测量误差
2. 工具变量条件：
    1. 相关性：与内生解释变量相关
    2. 外生性：与扰动项不相关
3. 二阶段最小二乘法（2SLS/TSLS）：
    1. 第一阶段，内生解释变量对工具变量回归
    2. 第二阶段，被解释变量对第一阶段回归的拟合值再做回归

    3. 阶条件：工具变量个数不少于内生解释变量个数
        1. 不可识别：工具变量个数少于内生解释变量个数
        2. 恰足识别：工具变量个数等于内生解释变量个数
        3. 过度识别：工具变量个数大于内生解释变量个数
4. 弱工具变量
5. 过度识别检验：Sargan统计量
6. 内生性豪斯曼检验：Hausman test

二值选择模型

1. logit
2. probit

面板数据

数据特征

分类：

- 短面板/微观面板：T 较小而 n 较大，较为常见
- 长面板/宏观面板：T 较大而 n 较小
- 动态面板：解释变量包含被解释变量的滞后值
- 静态面板：解释变量不包含被解释变量的滞后值
- 平衡面板：每个时期在样本中的个体完全一样
- 非平衡面板：每个时期在样本中的个体不完全一样

优点：

- 可以解决遗漏变量问题
- 提供更多个体动态行为的信息
- 样本容量较大

问题：

- 通常不满足独立同分布的假定，因为同一个体在不同期的扰动项一般存在自相关
- 数据收集成本较高

估计策略

极端的策略：

1. **混合回归**：直接把面板当成截面，要求样本中每个个体都拥有完全相同的回归方程。缺点是忽略了个体不可观测的异质性（也就是下面个体效应模型中的 u_i），它可以和解释变量相关从而导致估计不一致（也就是固定效应模型的情况）
2. **分别回归**：每个个体都估计一个独立的方程。缺点是忽略了个体的共性，而且可能没有足够大的样本容量（特别是短面板）

折中的策略：假定每个个体的回归方程拥有相同的斜率，但可以有不同的截距项。这种模型称为**个体效应模型**

$$y_{it} = \mathbf{x}_{it}^T\beta + \mathbf{z}_i^T\delta + u_i + \varepsilon_{it} \quad (i = 1, \cdots, n; t = 1, \cdots, T) \tag{28}$$

其中，

- $\mathbf{z}_i$ 为不随时间改变的个体特征，比如性别
- $\mathbf{x}_{it}$ 可以随个体和时间而改变
- 扰动项由 $(u_i + \varepsilon_{it})$ 两个部分组成，称为复合扰动项
    - u_i 为个体异质性的截距项，即**个体效应**
        - 如果 u_i 与某个解释变量相关，则进一步称为**固定效应模型**
        - 如果 u_i 和所有的解释变量 $(\mathbf{x}_{it}, \mathbf{z}_i)$ 均不相关，则进一步成为**随机效应模型**
    - ε_{it} 为随个体和时间改变的扰动项。一般假设其为独立同分布，且与 u_i 不相关

混合回归

所有个体都拥有完全一样的回归方程，也就没有了个体效应，即 $u_1 = u_2 = \cdots = u_n$，统一记为 α

$$y_{it} = \alpha + \mathbf{x}_{it}^T\beta + \mathbf{z}_i^T\delta + \varepsilon_{it} \tag{29}$$

其中 x_{it} 不含常数项。这时候就可以把所有数据放在一起当做横截面一样直接 OLS

由于面板数据的特点，这种情况下虽然可以假设不同个体之间的扰动项相互独立，但是同一个体在不同期的扰动项往往都有自相关。所以每个个体在不同时期的所有观测值构成一个“**聚类**”。同一聚类内部的观测值互相相关，不同聚类之间的观测值则不相关。对于聚类样本的 OLS 统计，需要用到**聚类稳健的标准误**。适用于 T 较大，n 较小的短面板。另外这种聚类稳健标准误也是异方差稳健的。

个体固定效应

组内估计量

因为个体效应 u_i 和某个解释变量相关，所以 OLS 不一致。需要想办法消掉这个 u_i 项。对给定个体 i，在原个体效应模型两边对时间取平均

$$\bar{y}_i = \bar{\mathbf{x}}_i^T \beta + \mathbf{z}_i^T \delta + u_i + \overline{\varepsilon_i} \tag{30}$$

用原方程减去上面的平均方程，可以得到模型的离差形式：

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^T \beta + (\varepsilon_{it} - \bar{\varepsilon}_i) \tag{31}$$

记为

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}^T \beta + \tilde{\varepsilon}_{it} \tag{32}$$

这样只要新的扰动项 $\tilde{\varepsilon}_{it}$ 与新的解释变量 $\tilde{\mathbf{x}}_{it}$ 不相关，就可以用 OLS 得到一致的估计 β。这也就是固定效应估计量，因为用到了每个个体的组内离差信息，所以也称为**组内估计量**。记为 $\hat{\beta}_{FE}$。

考虑到可能存在组内自相关，**也应该用每个个体为聚类的聚类稳健标准误**。

即便个体特征 u_i 与某个解释变量相关，只要用组内估计量就可以得到一致估计，这是面板数据的一大优势。但是，在做离差变换的过程中，同时还消掉了 $\mathbf{z}_i^T \delta$，也就无法估计不随时间变化的变量 δ，这是固定效应模型的一大缺点。

另外，为了保证新的扰动项和新的解释变量不相关。**还需要假定个体 i 满足严外生**，也就是说 $E(\varepsilon_{it} | \mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT}) = 0$，这是因为 $\bar{\mathbf{x}}_i$ 中包含了所有 $(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT})$ 的信息。这个假定就要比一般大样本要求的前定变量或者叫同期外生的假定要强。

最小二乘虚拟变量法 LSDV

前面说到 u_i 为个体异质性的截距项。对于 n 个个体的 n 个不同截距项，可以通过在 (28) 式中引入 ($n-1$) 个**个体**虚拟变量来体现（如果没有截距项，则引入 n 个虚拟变量）

$$y_{it} = \alpha + \mathbf{x}_{it}^T \beta + \mathbf{z}_i^T \delta + \sum_{i=2}^{n} \gamma_i D_i + \varepsilon_{it} \tag{33}$$

注意上式中有单独截距项 α ，不同于混合回归中的统一截距项。而是被遗漏虚拟变量 D_1 对应个体 1 的截距项

其中，如果为个体 2，则个体虚拟变量取值 $D_2 = 1$；否则 $D_2 = 0$。如此类推。从而个体 i ($i > 1$) 的截距项是 ($\alpha + \gamma_i$) ，**不只是** γ_i

LSDV 方法对 β 的估计结果和组内估计量是完全一样的。但是如果做完之后因为某些个体的虚拟变量不显著而删去了，那就不相同了。

LSDV 的好处是可以估计个体异质性 u_i；坏处是个体数量 n 很大时可能需要引入很多个虚拟变量

一阶差分法

在 (28) 式两边关于时间 t 进行一阶差分，同样可以消去个体效应 u_i（同时也消掉了 $\mathbf{z}_i^T\delta$）

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})^T\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}) \tag{34}$$

对上式使用 OLS 就得到了一阶差分估计量 $\hat{\beta}_{FD}$

现在只要扰动项的一阶差分 $(\varepsilon_{it} - \varepsilon_{i,t-1})$ 与解释变量的一阶差分 $(\mathbf{x}_{it} - \mathbf{x}_{i,t-1})$ 不相关即可得到一致估计，这个假定就比组内估计量的严外生要求要弱。这也是一阶差分法的主要优势。

但是，只有 $T=2$ 时 $\hat{\beta}_{FD} = \hat{\beta}_{FE}$。$T > 2$ 且 $\{\varepsilon_{it}\}$ 独立同分布，则还是 $\hat{\beta}_{FE}$ 更有效率。**所以实践中还是用组内估计量的多，用一阶差分法的少。**

时间固定效应

个体固定效应处理了不随时间变化但随个体而异的遗漏变量问题。但还可能有不随个体而变，只随时间变化的遗漏变量问题。为此，在 (28) 式中加入时间固定效应 λ_t

$$y_{it} = \mathbf{x}_{it}^T\beta + \mathbf{z}_i^T\delta + \lambda_t + u_i + \varepsilon_{it} \tag{35}$$

对于此方程，可以用 LSDV 法来估计。同样对每一个时期定义一个虚拟变量，然后把 $(T-1)$ 个**时间**虚拟变量包括在回归方程中（未被包括的就是基期）

$$y_{it} = \alpha + \mathbf{x}_{it}^T\beta + \mathbf{z}_i^T\delta + \sum_{i=2}^{T}\gamma_t D_t + u_i + \varepsilon_{it} \tag{36}$$

和个体固定效应的 LSDV 部分类似：常数项 α 为 D_1 所对应的第 1 期截距项，而第 t 期（$t>1$）截距项为 $(\alpha + \gamma_t)$

上面的方程既考虑了个体固定效应，又考虑了时间固定效应，所以称为“**双向固定效应**”；相对的只考虑个体固定效应的则称为“单向固定效应”。可以通过检验这些时间虚拟变量的联合显著性来判断是不是应该用双向固定效应模型。

T 比较大时为了节省参数，可以引入一个时间趋势项用来替代 $(T-1)$ 个时间虚拟变量

$$y_{it} = \mathbf{x}_{it}^T\beta + \mathbf{z}_i^T\delta + \gamma t + u_i + \varepsilon_{it} \tag{37}$$

这样做隐含的假定是每期的时间效应相等，每期都增加 γ。如果这个假定明显不成立，那还是不能偷懒，只能引入虚拟变量

随机效应

非平衡面板

豪斯曼检验

究竟用固定效应还是随机效应

平稳时间序列

1. 自回归
2. ARMA
3. VAR

单位根与协整

1.

显著性驱动的选题和处理

由 $Var(\hat{\beta}|X)=\sigma^2(X^TX)^{-1}$ 或者说 $Var(\hat{\beta}_k|X)=\dfrac{\sigma^2}{(1-R_k^2)S_k}$ 的公式可以知道，为了更可能的得到显著，应该选择符合以下性质的题目：

- 取对数更容易显著
- 可获得样本量尽可能多，X 的扰动越大越好，也就是说要找外生冲击 shock
- 多重共线性问题：X 内部各个变量（列）最好完全独立，越不相关越好

具体的统计量

直观上，如果未知参数的估计值离预期假想值比较“远”，则应倾向于拒绝原假设（H_0：未知参数等于预期假想值）。使用此原理的这类统计检验成为沃尔德检验（Wald test）

t 统计量

小样本

满足全部五个小样本 OLS 古典假定，并且原假设 $H_0:\beta_k=c$ 也成立的情况下

$$t_k=\frac{\hat{\beta}_k-c}{SE(\hat{\beta}_k)}\sim t(n-K) \tag{38}$$

一般的通用公式为

$$t \equiv \frac{\text{估计量 一 假想值}}{\text{估计量的标准误}} \tag{39}$$

大样本

$$t_k = \frac{\hat{\beta}_k - c}{SE_1(\hat{\beta}_k)} \xrightarrow{d} N(0,1) \tag{40}$$

其中 $SE_1(\hat{\beta}_k) = \sqrt{\frac{1}{n} A\hat{var}(\hat{\beta}_k)}$

服从正态分布而不再是 t 分布

F 统计量

小样本

原假设为检验多个线性约束条件，矩阵表示为 $H_0 = \mathbf{R}_{m*K}\beta_{K*1} = \mathbf{r}_{m*1}$；其中 $\mathbf{R}$ 要满行秩，也就是 $rank(\mathbf{R}) = m$，待检验线性约束条件不能有多余或者自相矛盾

根据 Wald test 原理，由于 $\hat{\beta}$ 是 β 的估计量，故如果 H_0 成立，那么 $\mathbf{R}\hat{\beta} - \mathbf{r}$ 应该比较接近 $\mathbf{0}$。这种接近程度用其二次型来衡量，比如说

$$(\mathbf{R}\hat{\beta} - \mathbf{r})^T [Var((\mathbf{R}\hat{\beta} - \mathbf{r}))]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) \tag{41}$$

其中,

$$\begin{aligned}
Var(\mathbf{R}\hat{\beta} - \mathbf{r}) &= Var(\mathbf{R}\hat{\beta}) && \text{(42)} \\
&= \mathbf{R} Var(\hat{\beta}) \mathbf{R}^T && \text{（夹心估计量）} \quad \text{(43)} \\
&= \sigma^2 \mathbf{R}(\mathbf{X^T X})^{-1} \mathbf{R}^T && \text{（因为} Var(\hat{\beta}) = \sigma^2 (\mathbf{X^T X})^{-1}\text{）} \quad \text{(44)}
\end{aligned}$$

从而在满足全部五个小样本 OLS 古典假定，并且原假设 $H_0 = \mathbf{R}\beta = \mathbf{r}$ 的情况下，有

$$F \equiv \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})^T [\mathbf{R}(\mathbf{X^T X})^{-1} \mathbf{R}^T]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r})/m}{s^2} \sim F(m, n-K) \tag{45}$$

另外，**在做假设检验时，如果接受原假设，则可将原假设作为约束条件，代入最小二乘法的最优化问题。**

从而 F 检验可以变为

$$\begin{gathered}
\min_{\hat{\beta}} \ SSR(\hat{\beta}) \qquad (46) \\
s.t. \ \mathbf{R}\hat{\beta} = \mathbf{r}
\end{gathered}$$

如果原假设 $H_0 : \mathbf{R}\beta = \mathbf{r}$ 正确，那么加上约束后不应使得残差平方和增大很多

记无约束回归的残差平方和为 SSR_0，而有约束回归的残差平方和为 SSR_1。这意味着，在 H_0 正确的情况下，$(SSR_0 - SSR_1)$ 不应很大。从而构造如下的 F 统计量

$$
\begin{aligned}
F &= \frac{(SSR_1 - SSR_0)/m}{SSR_0/(n-K)} \quad (47) \\
&= \frac{(R_0^2 - R_1^2)/m}{(1-R_0^2)/(n-K)} \quad (48)
\end{aligned}
$$

其中用了 $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}$ 来替换，注意第二个等号后的 R^2 都加了负号

大样本

假设统计量 $F \sim F(m,n)$，则当 $n \to \infty$ 时，$mF \xrightarrow{d} \chi^2(m)$