线性面板数据分析

司继春

2025年5月

面板数据

- 面板数据（panel data），又称为纵向数据（longitudinal data），指的是在不同的T个时间点对相同的N个体进行观测。
  - 比如，中国2010年到2017年共8年时间每个省份的GDP、人均收入等变量就形成了一个面板数据。
- 在这一章里，我们主要关心N大T小的短面板（short panel），即$N \to \infty$，而T是固定的。
- 此外，在本章中，我们同时假设所使用的面板数据为**平衡面板（balanced panel）**数据。
  - 如果在每一期，我们观察到了同样多的相同的个体，那么我们称之为平衡面板。
  - 而相反，在某些时期一些个体出于种种原因没有出现在数据中，则称之为**非平衡面板（unbalanced panel）**。
  - 非平衡面板主要是由于数据缺失或者调查时迭代更新样本造成的。如果个体数据缺失是随机的，不会对我们接下来的结果有本质影响，然而如果缺失是非随机的、系统性的，非平衡面板需要更加详细的讨论。

面板数据模型

- 在面板数据的分析中，与截面数据不同的是，我们可以允许不同时间、不同个体具有不同的截距项：

$$y_{it} = \alpha_{it} + x_{it}'\beta + u_{it}$$

  其中$i = 1, ..., N$代表个体，而$t = 1, ..., T$为时间，α_{it}为个体-时间可变的截距项，如果是平衡面板，那么我们将会有$N \times T$个观测。
- 在面板数据中，个体-时间可变的截距项α_{it}是相对截面数据中特别的设定，该截距项允许每个个体、每一期都有不同的截距项。
- 然而，该模型设定太过于一般化：由于α_{it}同时随着时间和个体变化，且是不可观测的，如果我们不做任何假定，以上模型将不可被识别。
- 实际上，按照以上的设定，那么α_{it}的个数为$N \times T$个，与样本量已经相等，所以无法使用现有样本直接估计出所有的α_{it}。

不可观测异质性

- 为了识别以上模型，一个非常常用的假设是假设α_{it}可以被分解为可加的两部分：

$$\alpha_{it} = \alpha_i + \lambda_t$$

得到

$$y_{it} = \alpha_i + \lambda_t + x'_{it}\beta + u_{it}$$

其中α_i为不随时间变化（time-invariant）的个体效应（individual effects）或者个体异质性（individual heterogeneity），而λ_t则代表了不同时间截距项的不同，也就是时间异质性。

不可观测异质性

- 其中，个体效应α_i度量了所有的个体（可观测与不可观测的、影响y的）不随时间变化而变化的特征，比如一个人的性别、出生地、出生年份，甚至可能不随时间变化的性格、能力等；或者对于企业而言，企业的成立时间、地理位置、不随时间变化的企业文化等，都可以被包含在α_i中。
- 而λ_t则度量了同一时间对于所有个体都相同的y的冲击
  - 比如宏观经济运行情况变量，可能对于所有企业有共同的影响，比如当GDP存在正向冲击时，所有企业都会受到大环境影响，而λ_t建模了这种共同冲击的影响。
  - 当然现实中不同个体对于共同冲击的反应可能是有不同的：交互固定效应

时间固定效应

- 对于λ_t，一种处理方法是直接设定t的函数形式，比如线性趋势：

$$\lambda_t = \lambda t$$

  或者二次趋势：

$$\lambda_t = \lambda_1 t + \lambda_2 t^2$$

  等。
- 然而这些设定都有很强的函数形式假定：对于任意的非线性的、偶发的冲击都无法很好的建模。

时间固定效应

- 由于我们在这里假设短面板数据，面临的是T固定但是$N \to \infty$的情况，因而λ_t通常比较容易处理，可以直接通过引入**时间固定效应（time fixed efftcts）**，即当前观测属于哪一期的虚拟变量来解决。
- 在引入了时间固定效应后，所有的不随个体i而变化的变量（比如国家的GDP、利率等）都与λ_t共线，从而不能再加入到回归方程中。
- 不失一般性，我们通常将时间固定效应λ_t看做是x_{it}中的一部分，因而我们可以将模型写为

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}$$

其中现在的x_{it}中包含着时间固定效应λ_t。

时间固定效应

值得注意的是，在面板数据中，时间固定效应一般是**必须**加入的，否则很有可能会因为变量都有趋势而发生“**伪回归**”（**spurious regression**）的情况，即回归系数并不是因为x和y之间的因果关系，而是共同的趋势或者冲击导致的。

伪回归

- 如果$x_{it} = f_{it} + \lambda_t, y_{it} = g_{it} + \lambda_t$，其中$f_{it} \perp\!\!\!\perp g_{it}$，而$\lambda_t$为同时影响$x$和$y$的共同的冲击，与$f_{it}, g_{it}$独立。
- 此时x和y之间并没有因果关系，但是由于受到了共同的冲击λ_t的影响，导致$\mathbb{C}\left(x_{it}, y_{it}\right) \neq 0$，从而$y_{it}$对$x_{it}$的回归系数也不为0。
- 但是由于$y_{it} \perp\!\!\!\perp x_{it}|\lambda_t$，所以一旦加入了时间固定效应，那么$x_{it} - \mathbb{L}\left(x_{it}|\lambda_t\right) = f_{it}, y_{it} - \mathbb{L}\left(y_{it}|\lambda_t\right) = g_{it}$，从而回归系数$\frac{\mathbb{C}(f_{it}, g_{it})}{\mathbb{V}(f_{it})} = 0$就能排除共同冲击的影响了。

面板的向量形式

- 接下来，在本章中，我们都以

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}$$

为基础进行讨论。

- 为了方便起见，我们通常记

$$X_i = \begin{bmatrix} x_{i1}' \\ \vdots \\ x_{iT}' \end{bmatrix}_{T\times K}, Y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}_{T\times 1}, U_i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}_{T\times 1}$$

从而上式可以写为

$$Y_i = \alpha_i + X_i\beta + U_i$$

由于T的固定的，从而以上X_i, Y_i, U_i的维数都是固定的。

面板的向量形式

通常，为了获得统计性质，我们会做如下假定:

独立同分布假定

假设$\{(Y_i, X_i, U_i), i = 1, ..., N\}$是独立同分布的。

- 注意在以上假设中，我们实际上假设的是不同个体之间独立同分布，而没有假定同一个个体i的不同期的独立性或者同分布性。
- 实际上，对于同一个个体i，如果$s \neq t$，x_{it}和x_{is}、u_{it}和u_{is}通常都是相关的，从而同一个个体的不同期通常是不独立的。
- 我们假设样本之间是独立同分布的，但是并不假设同一个个体的不同观测之间的独立性或者同分布性。

两种不同的模型

- 虽然对于时间固定效应λ_t的处理相对简单，然而，由于$N \to \infty$，对α_i的处理相对困难。
- 注意α_i也是一个随机变量，根据对α_i的假设不同，通常来讲有两类不同的模型可以使用：
  - 随机效应（random effects）模型
  - 固定效应（fixed effects）模型
- 两者的关键区别在于α_i与解释变量x_{it}之间相关性假设。在介绍两种方法之前，我们首先介绍使用传统的最小二乘法及其所需要的假设。

两种不同的模型

- 对于面板数据，如果忽略其面板结构，我们可以直接通过OLS进行估计。
- 然而，注意到由于α_i不可观测，从而模型可以写为

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it} \triangleq x_{it}'\beta + v_{it}$$

  其中v_{it}包含所有不可观测的因素，包括误差项u_{it}以及个体异质性α_i。
- 针对以上的设定，我们首先可以忽略面板结构，将每个个体的每一期都当做是一个独立的个体，从而，我们将$N \times T$个观测视为独立的个体。

外生性假设

根据之前OLS一致性的讨论，为了使得一致性成立，我们需要假设$\mathbb{E}\left(v_{it}|x_{it}\right)=0$，而为了保证$\mathbb{E}\left(v_{it}|x_{it}\right)=0$成立，我们通常将其分解为两个单独的假设：

当期外生性，contemporaneous exogeneity

假设对于所有的i,t，有$\mathbb{E}\left(u_{it}|x_{it}\right)=0$。

以及：

个体异质性的外生

假设对于所有的i，有$\mathbb{E}\left(\alpha_i|X_i\right)=0$。

外生性假设

- 注意以上仅仅假设了当期的u_{it}和当期的x_{it}不相关
- 但是由于不可观测的个体异质性α_i出现在了每一期的方程中，因而需要α_i和所有期的x_{it}都不相关，可以简写为α_i和X_i不相关。
- 由于本章中主要考虑个体异质性α_i的问题，因而我们暂且忽略u_{it}与x_{it}的相关性。
- 当然在实证中，如果u_{it}与x_{it}相关，那么这种外生性的违背是需要重点考虑的。

POLS

- 注意在这里虽然我们将解释变量x_{it}的下角标写成it，但是在实际应用时，仍然需要加入那些个体i的随时间不变的特征$\tilde{x}_i$
- 由于在模型中加入了时间固定效应，所以所有的不随个体变化、只随时间变化的变量都可以忽略，然而我们没有加入个体i的固定效应，所以那些不随时间变化的个体特征（如性别等）仍然不与其他任何解释变量共线，需要被加入到模型中。
- 否则，这些不随时间变化的个体特征会进入到α_i从而误差项v_{it}中，使其更容易与x_{it}相关，从而外生性更难以满足。

POLS

- 在以上假设条件下，我们可以将v_{it}视为新的误差项，此时我们可以简单的使用OLS对其进行估计。
- 若以上假设成立：

$$\mathbb{E}\left(x_{it}v_{it}\right)=\mathbb{E}\left[\mathbb{E}\left(x_{it}v_{it}\right)|x_{it}\right]=\mathbb{E}\left[x_{it}\mathbb{E}\left(\alpha_i+u_{it}|x_{it}\right)\right]=0$$

从而$\mathbb{E}\left(v_{it}x_{it}\right)=0$，由此得到矩条件：

$$\mathbb{E}\left[x_{it}\left(y_{it}-x_{it}'\beta\right)\right]=0$$

得到

$$\beta=\left[\mathbb{E}\left(x_{it}x_{it}'\right)\right]^{-1}\mathbb{E}\left(x_{it}y_{it}\right)$$

POLS

- 从而矩估计：

$$
\begin{aligned}
\hat{\beta}^{\text{POLS}} &= \left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}x_{it}')\right]^{-1}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}y_{it})\right] \\
&= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}x_{it}')\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}y_{it})\right] \\
&= \left(\sum_{i=1}^{N}X_i'X_i\right)^{-1}\left(\sum_{i=1}^{N}X_i'Y_i\right)
\end{aligned}
$$

以上估计方法实际上就是使用OLS直接进行估计，因而也被称为**混合最小二乘法**（**pooled ordinary least squares，POLS**）。

POLS

如果记

$$X=\begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}_{NT\times K}, Y=\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}_{NT\times 1}$$

那么POLS估计量可以写为:

$$\hat{\beta}^{\text{POLS}}=\left(X'X\right)^{-1}X'Y$$

从而获得POLS估计量的一个重要前提是$(X'X)^{-1}$可逆。据此，我们可以引入如下识别条件:

识别条件

假设$\mathbb{E}\left(X_i'X_i\right)$可逆。

POLS统计性质

- 如果记$V_i = \alpha_i + U_i$，那么回归式可以写为$Y_i = X_i\beta + V_i$
- 将其带入到混合最小二乘的估计式中，有：

$$
\begin{aligned}
\hat{\beta}^{\text{POLS}} &= \left(\sum_{i=1}^{N} X_i'X_i\right)^{-1}\left[\sum_{i=1}^{N} X_i'\left(X_i\beta + V_i\right)\right] \\
&= \beta + \left(\sum_{i=1}^{N} X_i'X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'V_i\right) \\
&= \beta + \left(\sum_{i=1}^{N} X_i'X_i\right)^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}v_{it}\right)\right]
\end{aligned}
$$

POLS统计性质

- 从而在外生性假设以及一定可积性条件下，根据大数定律，有：

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}v_{it}) \xrightarrow{p} \mathbb{E}\left(\sum_{t=1}^{T}(x_{it}v_{it})\right) = \sum_{t=1}^{T}\mathbb{E}(x_{it}v_{it}) = 0$$

进而易得，POLS估计量为一致估计量。

- 进一步，由于

$$\begin{cases} \frac{1}{N}\sum_{i=1}^{N} X_i'X_i \xrightarrow{p} \mathbb{E}(X_i'X_i) \\ \sqrt{N}\frac{1}{N}\sum_{i=1}^{N} X_i'V_i \overset{a}{\sim} \mathcal{N}(0, \mathbb{V}(X_i'V_i)) \end{cases}$$

而其中$\mathbb{V}(X_i'V_i) = \mathbb{E}(X_i'V_iV_i'X_i)$，从而

$$\sqrt{N}\left(\hat{\beta}^{\text{POLS}} - \beta\right) \overset{a}{\sim} \mathcal{N}\left(0, \left[\mathbb{E}(X_i'X_i)\right]^{-1}\mathbb{E}(X_i'V_iV_i'X_i)\left[\mathbb{E}(X_i'X_i)\right]^{-1}\right)$$

POLS的标准误

- 为了估计混合最小二乘估计量的标准误，类比异方差稳健标准误的计算方法。
- 记$\hat{V}_i = Y_i - X_i \hat{\beta}^{\text{POLS}}$为个体$i$的残差向量，那么使用插入法：

$$\widehat{\mathbb{V}\left(\hat{\beta}^{\text{POLS}}\right)} = (X'X)^{-1} \left[\sum_{i=1}^{N} \left(X_i' \hat{V}_i \hat{V}_i' X_i\right)\right] (X'X)^{-1}$$

以上就得到了POLS估计系数的标准误的一致估计。注意其中：

$$\begin{aligned}
\sum_{i=1}^{N} \left(X_i' \hat{V}_i \hat{V}_i' X_i\right) &= \sum_{i=1}^{N} \left(\begin{bmatrix} x_{i1} & \cdots & x_{iT} \end{bmatrix} \begin{bmatrix} \hat{v}_{i1} \\ \vdots \\ \hat{v}_{iT} \end{bmatrix} \begin{bmatrix} \hat{v}_{i1} & \cdots & \hat{v}_{iT} \end{bmatrix} \begin{bmatrix} x_{i1}' \\ \vdots \\ x_{iT}' \end{bmatrix} \right) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{v}_{it} \hat{v}_{is} x_{it} x_{is}'\right) \\
&= \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\hat{v}_{it}^2 x_{it} x_{it}'\right) + \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1, s \neq t}^{T} \left(\hat{v}_{it} \hat{v}_{is} x_{it} x_{is}'\right)
\end{aligned}$$

- 聚类标准误：至少聚类到个体i层面

POLS实例

俄罗斯大选

在之前的例子中，我们展示了Enikolopov, Petrova和Zhuravskaya（2011）的截面数据分析结果，除此之外文章中还使用了如下的面板数据设定：

$$\text{vote}_{st} = \beta_0 + \beta_1 \text{NTV}_{st} + x'_{st}\beta + \lambda_t + \alpha_i + \epsilon_{st}$$

其中$t = 1995, 1999$，即一个两年的面板数据，s为选区，x_{st}为其他控制变量。如果使用POLS对以上模型进行回归，那么可以使用如下代码：

POLS实例

俄罗斯大选

```
use datasets/NTV_Aggregate_Data_reshaped.dta
// 年1995并没有成立NTV
gen Watch_probit_p=0
replace Watch_probit_p=Watch_probit if year==1999
// 回归POLS
reg Votes_SPS_ Watch_probit_p i.region i.year if year!=2003,
    cluster(tik_id)
// reghdfe
reghefe Votes_SPS_ Watch_probit_p if year!=2003, a(i.region i.year
    ) cluster(tik_id)
```

随机效应模型

- 虽然我们可以使用混合最小二乘得到估计，但是由于α_i的存在，误差项V_i的协方差矩阵不会是一个同方差、无自相关的理想情况
- 而这种理想情况是BLUE所需要的，这也就意味着有效性的缺失。
- 为了达到有效性，我们可以使用误差项V_i的协方差结构进行**广义最小二乘**（generalized least squares，GLS）。

随机效应的假设

协方差结构和外生性假设

假设:

1. α_i具有同方差，即$\mathbb{V}\left(\alpha_i | X_i\right) = \sigma_\alpha^2$

2. u_{it}具有同方差且无自相关，即假设$\mathbb{V}\left(U_i | X_i\right) = \sigma_u^2 \cdot I_T$。

3.$\mathbb{E}\left(\alpha_i u_{it} | X_i\right) = 0$

随机效应模型的协方差结构

- 在这里，我们将α_i视为一个均值独立于x_i的随机变量，因而我们通常将该设定成为**随机效应（random effects）**模型。
- 根据以上假设，我们可以计算：

$$
\begin{aligned}
\Omega &= \mathbb{V}\left(V_i|X_i\right) = \mathbb{V}\left(\alpha_i + U_i|X_i\right) \\
&= \sigma_u^2 I_T + \sigma_\alpha^2 \iota_T \iota_T' \\
&= \begin{bmatrix} \sigma_\alpha^2+\sigma_u^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2+\sigma_u^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2+\sigma_u^2 \end{bmatrix}
\end{aligned} \tag{1}
$$

因为有了α_i的存在，$\mathbb{C}\left(v_{it}, v_{is}\right) \neq 0$，从而存在了自相关

随机效应模型的估计

- 如果我们知道Ω，就可以得到最有效的估计量。
- 如果我们在方程的两边同时乘以$\Omega^{-1/2}$，得到：

$$\Omega^{-1/2}Y_i = \Omega^{-1/2}X_i\beta + \Omega^{-1/2}V_i$$

如果记

$$Y_i^* = \Omega^{-1/2}Y_i, X_i^* = \Omega^{-1/2}X_i, V_i^* = \Omega^{-1/2}V_i$$

那么以上方程可以写为：

$$Y_i^* = X_i^*\beta + V_i^*$$

其中Y_i^*为$T \times 1$的向量，而X_i^*为$T \times K$的矩阵。

随机效应模型的估计

- 此外，有：

$$\mathbb{V}\left(V_i^*\right) = \mathbb{V}\left(\Omega^{-1/2}V_i\right) = \Omega^{-1/2}\Omega\Omega^{-1/2} = I$$

因而我们可以直接使用Y_i^*对X_i^*做OLS，得到：

$$\begin{aligned}
\hat{\beta}^{\text{RE}} = \hat{\beta}^{\text{GLS}} &= \left(\sum_{i=1}^{N} X_i^{*\prime} X_i^*\right)^{-1} \left(\sum_{i=1}^{N} X_i^{*\prime} Y_i^*\right) \\
&= \left(\sum_{i=1}^{N} X_i'\Omega^{-1}X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i'\Omega^{-1}Y_i\right) \\
&= \left[X'\left(I_N \otimes \Omega^{-1}\right)X\right]^{-1}\left[X'\left(I_N \otimes \Omega^{-1}\right)Y\right]
\end{aligned}$$

即随机效应估计量（random effects estimator）。

随机效应的假设

- 值得注意的是，虽然在当期外生性的条件下，混合最小二乘是一致的，但是这并不代表随机效应估计量一定一致。
- 注意到

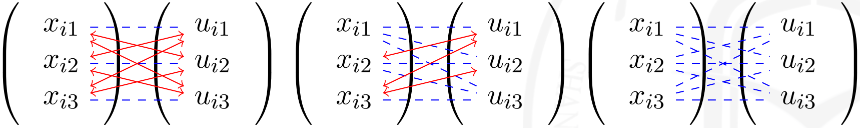
$$\hat{\beta}^{\text{GLS}} = \left(\sum_{i=1}^{N} X_i'\Omega^{-1}X_i\right)^{-1}\left[\sum_{i=1}^{N} X_i'\Omega^{-1}\left(X_i\beta + V_i\right)\right]$$
$$= \beta + \left(\sum_{i=1}^{N} X_i'\Omega^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'\Omega^{-1}V_i\right)$$

其中

$$\frac{1}{N}\sum_{i=1}^{N} X_i'\Omega^{-1}V_i \xrightarrow{p} \mathbb{E}\left(X_i'\Omega^{-1}V_i\right)$$

- 然而，如果Ω不是对角矩阵（从而Ω^{-1}也不是对角阵），在$X_i'\Omega^{-1}V_i$中就必然存在着$x_{is}u_{it}, s \neq t$
- 而当期外生性假设仅仅要求当期不相关，即$\mathbb{E}\left(x_{it}u_{it}\right) = 0$，而没有对$\mathbb{E}\left(x_{is}u_{it}\right)$做任何假定
- 因而如果仅仅假设当期不相关，可能会存在着不同期的x和u之间的相关性，此时$\mathbb{E}\left(X_i'\Omega^{-1}V_i\right) = 0$可能不成立。

随机效应的假设



(a) 当期外生性 (b) 序贯外生性 (c) 严格外生性

序贯外生性

- 比当期外生性更加严格一点的是序贯外生性：

序贯外生性，sequential exogeneity

假设对于所有的i, t，有$\mathbb{E}\left(u_{it} | x_{it}, x_{i,t-1}, \ldots, x_{i1}\right) = 0$。

- 上图(b)展示了序贯外生性。在当期外生性的基础上，序贯外生性不仅要求当期的u_{it}与x_{it}不相关，而且还额外要求u_{it}和过去的所有x_{it}（$x_{it}, x_{i,t-1}, \ldots, x_{i1}$）都不相关。
- 注意此时我们允许u_{it}和未来的x（$x_{i,t+1}, \ldots, x_{iT}$）相关。
- 此时，看起来好像x_{it}是由滞后的误差项（$u_{i,t-1}, \ldots, u_{1}$）所“决定”的，我们称这些x为前定变量（**predetermined variables**）。

严格外生性

- 然而以上序贯外生性还是无法保证$\mathbb{E}\left(X_i'\Omega^{-1}V_i\right)=0$成立，实际上为了保证该条件成立，我们需要更严格的外生性，即：

严格外生性，strict exogeneity

假设对于所有的i,t，有$\mathbb{E}\left(u_{it}|X_i\right)=0$。

- 上图(c)展示了严格外生性，即需要假设所有期的x和所有期的u都不相关，这是一个最严格的外生性假设。
- 在严格外生性的条件下，我们可以得到$\mathbb{E}\left(X_i'\Omega^{-1}V_i\right)=0$成立，从而GLS估计量或者随机效应估计量$\hat{\beta}^{\text{GLS}}$是$\beta$的一致估计。

严格外生性

- 以上使用了条件期望定义严格外生性，实际上在这里我们只需要一个稍微弱一点的假设：

严格外生性2

假设

$$\mathbb{E}\left(u_i \otimes X_i\right) = \mathbb{E}\begin{bmatrix} u_{i1}X_i \\ \vdots \\ u_{iT}X_i \end{bmatrix}_{T^2 \times K} = \mathbb{E}\begin{bmatrix} u_{i1}x_{i1}' \\ \vdots \\ u_{i1}x_{iT}' \\ \vdots \\ u_{iT}x_{i1}' \\ \vdots \\ u_{iT}x_{iT}' \end{bmatrix} = 0$$

POLS v.s. RE

经常会看到，即使面板数据，文献中使用随机效应模型的也越来越少，与POLS相比，其在假设上有一点细微差别：

- 对α_i的假设上：两者都需要假设$\mathbb{E}\left(\alpha_i|x_i\right)=0$
- 对u_{it}的假设上：
  - POLS只需要假设$\mathbb{E}\left(u_{it}|x_{it},\alpha_i\right)=0$（当期外生性，**contemporaneous exogeneity**）
  - RE需要假设$\mathbb{E}\left(u_{it}|X_i,\alpha_i\right)=0, X_i=\left[x'_{i1},...,x'_{iT}\right]'$（严格外生性，**strict exogeneity**）

标准误的问题

- 渐进分布：

$$\sqrt{N}\left(\hat{\beta}^{\mathrm{GLS}}-\hat{\beta}\right) \overset{a}{\sim} \mathcal{N}\left(0, \left[\mathbb{E}\left(X_i'\Omega^{-1}X_i\right)\right]^{-1}\right)$$

  仍然依赖于式(1)的协方差结构假设。
- 实践中：仍然推荐使用聚类稳健标准误，至少聚类到个体层面。

随机效应实例

俄罗斯大选

在俄罗斯大选的例子中，如果需要使用随机效应估计量，可以使用：

```
xtreg Votes_SPS_ Watch_probit_p i.region i.year if year!=2003, re
    cluster(tik_id)
```

其中xtreg命令为面板数据回归命令，而re选项则指定使用随机效应模型进行估计，此外，标准误被聚类到个体层次。

- 在xtreg命令中，robust选项与cluster(panel_id)是等价的，由于该例该面板本身为选区层面面板，此时使用cluster(tik_id)与直接加入robust选项是等价的。
- 但是在其他命令如reg、reghdfe等并没有这一设定，仍然需要手动使用cluster()进行聚类。

一阶差分方法

- 为了处理可能的α_i与X_i之间的相关性，一个更加聪明的办法是想办法在回归方程中将α_i消掉，从而可以完全无需对α_i和X_i之间的关系进行建模。
- 将α_i消掉最简单的做法是直接对回归方程式

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}$$

进行一阶差分（first difference），从而得到

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

其中$\Delta y_{it} = y_{it} - y_{i,t-1}$为一阶差分，其他类似。
- 由于α_i是不随时间变化而变化的，所以经过一阶差分之后，自然消掉了α_i，从而也就无需考虑α_i与x_{it}之间的相关性。
- 当然，解释变量中不随时间变化的所有特征也被一阶差分所消掉，所以使用该方法时，x_{it}中将不能包含不随时间而变化的解释变量（也不包含常数项）。

一阶差分方法

- 为了使得以上方程可以通过OLS直接估计，需要假定：

$$\mathbb{E}\left(\Delta u_{it} \Delta x_{it}\right) = \mathbb{E}\left[\left(u_{it} - u_{i,t-1}\right)\left(x_{it} - x_{i,t-1}\right)\right] = 0$$

- 因而序贯外生性等假设并不足以保证OLS估计量的一致性，至少需要假设相邻两期的u_{it}与x_{it}之间不相关才可以，当然，严格外生性足以保证，是更强的假设。

一阶差分的应用

美国媒体与选举

Gentzkow, Shapiro and Sinkinson(2011)研究了美国各县的报纸数量对美国选举投票率的影响。由于考虑到数据中冲击具有较强的持续性影响，从而选择了一阶差分模型，其模型设定为:

$$y_{ct} = \beta \times n_{ct} + x'_{ct}\delta + \alpha_c + \gamma_{st} + \epsilon_{ct}$$

其中y_{ct}为c县t年的总统选举投票率，n_{ct}为报纸数量，α_c为县固定效应，而γ_{ct}为州×年份固定效应。经过一阶差分后，得到:

$$\Delta y_{ct} = \beta \times \Delta n_{ct} + \Delta x'_{ct}\delta + \Delta\gamma_{ct} + \Delta\epsilon_{ct}$$

经过一阶差分后使用线性回归即可。（linear_panel_fd_np.do）

LSDV

- 然而，注意到如果将α_i视为虚拟变量，x_i为K维随时间变化的随机向量（从而不包含常数项），那么我们将有$N+K$个系数需要估计，观测数总共有$N \times T$个
- 一般而言$N \times T > N + K$，因而估计是可以进行的。
- 然而此时，自变量的维数随着样本量的增加而线性增加，一般而言是无法获得一致估计量的。在统计学中，这个问题通常被称为**伴随参数问题（incidental parameters problem）**。
- 不过幸运的是，在简单的线性模型中，以上LSDV估计量的确可以得到β的一致估计。我们可以使用分步回归证明这一点。

LSDV

- 注意到回归方程式中将α_i视为个体固定效应：

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}$$

如果使用最小二乘回归，等价于：

1. 使用将x_{it}对N个虚拟变量$\alpha_i, i = 1, ..., N$做回归，得到残差。根据之前的结论，该残差为$\ddot{x}_{it} = x_{it} - \bar{x}_i$其中$\bar{x}_i$为第$i$个个体所有时间段的$x$的平均值：$\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$以上过程即每个观测减去个体的平均，也成为去平均（demean）；
2. 使用y_{it}对N个虚拟变量$\alpha_i, i = 1, ..., N$做回归，得到残差。根据之前的结论，该残差为：$\ddot{y}_{it} = y_{it} - \bar{y}_i$即被解释变量去平均；
3. 使用OLS进行估计：

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

固定效应

- 实际上，以上估计量也可以由如下步骤得到：
  1. 对于回归方程式$y_{it}=\alpha_i+x_{it}'\beta+u_{it}$按照时间进行平均，得到：
  $$\bar{y}_i=\alpha_i+\bar{x}_i'\beta+\bar{u}_i$$
  2. 相减，得到：
  $$y_{it}-\bar{y}_i=(x_{it}-\bar{x}_i)'\beta+(u_{it}-\bar{u}_i)$$
  从而减去了个体异质性α_i
  3. 使用$\ddot{y}_{it}=y_{it}-\bar{y}_i$对$\ddot{x}_{it}=x_{it}-\bar{x}_i$做回归，得到估计量：
  $$\hat{\beta}^{\mathrm{FE}}=\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{y}_{it}\right)$$
- 传统上，在定义以上估计量时，通常将α_i视为固定的、不随机的参数，因而也被成为**固定效应**（**fixed effects**）模型。

固定效应

- 可见，以上两种方法：通过加入个体固定效应得到的LSDV估计量$\hat{\beta}^{\mathrm{LSDV}}$以及通过去平均得到的固定效应估计量$\hat{\beta}^{\mathrm{FE}}$是完全等价的。
- 实际上，根据我们之前的结论，加入个体固定效应意味着所有的比较都在组内（个体内部的不同时间之间）进行，因而以上估计量又被成为**组内估计量（within-group estimator）**。

固定效应统计性质

- 记$M_0^T=I-\frac{1}{T}\iota_T\iota_T'$，那么：

$$\begin{aligned}\hat{\beta}^{\text{FE}}&=\left(\sum_{i=1}^{N}\ddot{X}_i'\ddot{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}_i'\ddot{Y}_i\right)\\&=\left(\sum_{i=1}^{N}X_i'M_0^TX_i\right)^{-1}\left(\sum_{i=1}^{N}X_i'M_0^TY_i\right)\\&=\beta+\left(\sum_{i=1}^{N}X_i'M_0^TX_i\right)^{-1}\left(\sum_{i=1}^{N}X_i'M_0^TU_i\right)\end{aligned}$$

- 由于M_0^T不是对角阵，因而此时，我们需要**严格外生性**此时，有：

$$\frac{1}{N}\sum_{i=1}^{N}X_i'M_0^TU_i\xrightarrow{p}\mathbb{E}\left(X_i'M_0^TU_i\right)=0$$

从而一致性满足

固定效应的应用

俄国选举与NTV

在上面俄国选举的例子中，作者还额外使用1995年的选举结果进行了比对，其模型设定如下：

$$\text{vote}_{st} = \beta_0 + \beta_1 \text{NTV}_{st} + \delta_t + \alpha_s + \epsilon_{st}$$

由于以上只有两期，实际上以上做法与一阶差分是等价的。（linear_panel_fe_ntv.do）

可用的variation

- 组内估计量：估计
$$\ddot{y}_{it} = \ddot{x}_{it}'\beta + \ddot{u}_{it}$$
得到$\hat{\beta}^{\mathrm{WG}}$
  - 等价于LSDV，因而只包含了同一个体不同时间的比较。
- 组间估计量：估计
$$\bar{y}_i = \bar{x}_i'\beta + \bar{u}_i$$
得到$\hat{\beta}^{\mathrm{BG}}$
  - 只比较了不同个体的均值。
- 固定效应为组内估计量，而POLS、RE为两者的组合。

不同估计量之间的关系

$$\begin{array}{ccccccc}
\hat{\beta}^{\text{WG}} & = & \hat{\beta}^{\text{LSDV}} & = & \hat{\beta}^{\text{FE}} & \xleftarrow{\text{GLS}} & \hat{\beta}^{\text{FD}} \\
& \searrow & & & \wr\wr\; T\hat{\sigma}_{\alpha}^{2} \to \infty & & \\
& & \hat{\beta}^{\text{POLS}} & \xrightarrow{\text{GLS}} & \hat{\beta}^{\text{RE}} & & \\
& \nearrow & & & & & \\
\hat{\beta}^{\text{BG}} & & & & & &
\end{array}$$

POLS v.s. RE

- 两者都需要假设$\mathbb{E}\left(\alpha_i|x_i\right)=0$，同时使用了组内和组间信息
  - 在x_{it}中可以加入，也最好加入一些随时间不变的变量，以使$\mathbb{E}\left(\alpha_i|x_i\right)=0$更可能满足。
- 然而POLS仅仅需要当期外生性，而RE需要严格外生性。
  - No free lunch：RE即POLS经过GLS获得，理论上最有效，但是需要更严格的假设。

FE v.s. POLS

- 从假设强弱来看：
  - POLS需假设$\mathbb{E}\left(\alpha_i|x_i\right)=0$，但是只需要当期外生性
  - FE不需要假设$\mathbb{E}\left(\alpha_i|x_i\right)=0$，但是需要$U_i$的严格外生性
- 在POLS中往往也需要加入一些固定效应，区别在于：
  - FE需要加入个体固定效应
  - POLS往往需要加入上一级的固定效应
  - 比如在俄罗斯大选的例子中，加入地区v.s.选区的固定效应

FE v.s. RE

- 从假设强弱来看：
  - POLS需假设$\mathbb{E}\left(\alpha_i|x_i\right)=0$
  - FE不需要假设$\mathbb{E}\left(\alpha_i|x_i\right)=0$
  - 两者都需要U_i的严格外生性
- 两者都是基于某种“去平均”的计算方法：
  - FE：$\ddot{h}_{it}=h_{it}-\bar{h}_i, h\in\{y,x,u\}$
  - RE：$\check{h}_{it}=h_{it}-\left(1-\sqrt{\frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2+T\hat{\sigma}_\alpha^2}}\right)\bar{h}_i, h\in\{y,x,u\}$
  - 随着$T\hat{\sigma}_\alpha^2\to\infty$，$\text{FE}\approx\text{RE}$

FE v.s. RE：检验

- 固定效应与随机效应的关键区别：$\mathbb{E}(\alpha_i|x_i) = 0$是否成立：

	$\mathbb{E}(\alpha_i|x_i) = 0$	$\mathbb{E}(\alpha_i|x_i) \neq 0$
$\hat{\beta}^{\text{RE}}$	一致，最有效	不一致
$\hat{\beta}^{\text{FE}}$	一致	一致

- 理论上：如果$\mathbb{E}(\alpha_i|x_i) = 0$成立优先使用随机效应。
  - 实际上，如果该检验不显著，随机效应和固定效应的估计应该区别不大
- 检验：Hausman检验
- 实践中：固定效应最稳健，应优先考虑。
- 实践中：xtreg命令、reghdfe命令

FE v.s. FD

- 从假设强弱来看：
  - 两者均不需要假设$\mathbb{E}\left(\alpha_i | x_i\right) = 0$
  - FE需要U_i的严格外生性，FD仅仅需要相邻两期不相关
- 从有效性的角度：
  - 如果假设$\mathbb{V}\left(U_i\right) = \sigma_u^2 I$，则FE是最有效的，FD经过GLS后与FE等价
  - 如果假设$u_{it} = \rho u_{i,t-1} + e_{it}$，即单位根过程，FD是最有效的
  - 例：Gentzkow, Shapiro and Sinkinson（2011）

面板数据：标准误的问题

线性面板数据的标准误与OLS的标准误一样，同样存在Cluster的问题：

- 至少要Cluster到个体层面
    - xtreg命令加robust选项即可
    - 或者xtreg, cl(id)
- 当然，可以进行更高程度的Cluster
- 特别是使用reg、reghdfe等手动作固定效应时，一定不要忘了加cluster