

一、对几种面板数据模型的理解

1 混合效应模型 pooled model

就是所有的省份，都是相同，即同一个方程，截距项和斜率项都相同

$$y_{it}=c+bx_{it}+\varepsilon_{it} \quad c \text{ 与 } b \text{ 都是常数}$$

2 固定效应模型 fixed-effect model 和随机效应模型 random-effects model

就是所有省份，既有相同的部分，即斜率项都相同；也有不同的部分，即截距项不同。

2.1 固定效应模型 fixed-effect model

$$y_{it}=a_i+bx_{it}+\varepsilon_{it} \quad \text{cov}(c_i, x_{it}) \neq 0$$

固定效应方程隐含着跨组差异可以用常数项的不同刻画。每个 a_i 都被视为未知的待估参数。 x_{it} 中任何不随时间推移而变化的变量都会模拟因个体而已的常数项。

2.2 随机效应模型 random-effects model

$$y_{it}=a+u_i+bx_{it}+\varepsilon_{it} \quad \text{cov}(a+u_i, x_{it})=0$$

a 是一个常数项，是不可观察差异性的均值， u_i 为第 i 个观察的随机差异性，不随时间变化。

3 变系数模型 Variable Coefficient Models (变系数也分固定效应和随机效应)

每一个组，都采用一个方程进行估计。就是所有省份的线性回归方程的截距项和斜率项都不相同。

$$y_{it}=u_i+b_ix_{it}+\varepsilon_{it}$$

1. 混合估计模型就是各个截面估计方程的截距和斜率项都一样，也就是说回归方程估计结果在截距项和斜率项上是一样的。如果是考察各个省份，历年的收入对消费影响。则各个省份的回归方程就完全相同，无论是截距，还是斜率。

2. 随机效应模型和固定效应模型在斜率项都是相同的，都是截距项不同。区别在于截距项和自变量是否相关，不相关选择随机效应模型，相关选择固定效应模型。则说明各个省份的回归方程，斜率相同，差别的是截距项，即平移项。

3. 变系数模型，就是无论是截距项，还是系数项，对于不同省份，每个省份都有一个回归方程，都一个最适合自己的回归方程，完全不管整体。每个省份的回归方程与其他省份的，无论在斜率上，还是截距上都不相同。

总之，从混合估计模型，到变截距模型，再到变系数模型，考察省份是从完全服从整体和没有个性（回归方程是从整体角度而定的和估计的，是一刀切的，是完全没有差异性和个性的，完全牺牲自我），到随心所欲和完全个性化（每个省份都有一个最适合自己的回归方程）。即从完全无个性而言到完全有个性。

二、一个做医学哥们在固定效用模型和随机效用模型选择中的甄别方法

关于随机效应模型及固定效应模型的选择，一贯做法是两个模型都分析，看结果是否一致。如果一致且异质性较小或无，则选择固定效应模型。

如果结果不一致且异质性较大，则选择随机效应模型，并进行亚组分析寻找异质性的来源，并且下结论应比较保守。

好几篇 meta-analysis 在方法学部分都说：“All pooled outcome measures were determined using random-effects models”、“All pooled outcome measures were determined using random-effects models as described by DerSimonian and Laird”。

为什么都直接用随机效应模型却不用固定效应模型？是因为考虑 RCTs 异质性大，所以直接用随机效应模型吗？

1. 就是根据 I^2 值来决定模型的使用，大部分认为 $>50\%$ ，存在异质性，使用随机效应模型， $\leq 50\%$ ，用固定，有了异质性，通过敏感性分析，或者亚组分析，去探求异质性的来源，但是这两者都是定性的，不一定能找到，即使你做了，研究数目多的话，可以做个 meta 回归来找异质性的来源

2. 在任何情况下都使用随机效应模型，因为如果异质性很小，那么随即和固定效应模型最终合并结果不会有很大差别，当异质性很大时，就只能使用随机效应模型，所以可以说，在任何情况下都使用随机效应模型

3. 还有一种，看 P 值，一般推荐 P 的界值是 0.1，但现在大部分使用 0.05，就是说 $P > 0.05$ ，用固定， ≤ 0.05 用随机

其实个人偏向于第三种，因为 P 值可以看出来有没有异质性， I^2 是定量描述一致性的大小

本来随机效应的假设就是我们的样本从一个很大的母体抽取，所以大家的期望（均值）相同；如果我们的样本几乎是全部母体了，我们就不能说个体的差异是随机的，所以固定效应比较好；这是从模型的设定角度说的。但是随机效应模型有一个致命的硬伤，就是假设 $cov(x, u_i) = 0$ ，而固定效应不要求这个假设，Hausman 检验所做的工作就是检验一下这个假设对随机效应模型来说是不是成立，如果不成立，随即效应模型的估计是有偏的，即使采用 B-P 的 LM 检验表明存在随机效应，你也没有办法用了。

总结：检验固定效应是否显著，采用 F 检验（对比模型是 pooled）；检验随机效应是否显著，采用 LM 检验（对比模型也是 pooled）；检验固定和随机哪个更适用，采用 Hausman 检验（对比 fe 和 be）。

1 用 eviews 可以检验面板数据适用于混合估计法还是固定效应法

2 然后再进行豪斯曼检验，确定是用固定效应模型还是随机效应模型

三、是选择固定效应模型，还是随机效应模型的 Hausman test

Hausman test 是为了区分是选择固定效应模型，还是选择随机效应模型的计算。并且 Hausman test 是针对随机效应模型进行的检验，原假设是接受随机效应模型。

A Hausman test 说明一个有效的估计与它和一个非有效的估计之差的协方差是 0。即 $\text{Cov}(b-b, b) = \text{Cov}(b, b) - \text{var}(b) = 0$

B 原假设是随机效应模型有效，备选假设是固定效应模型有效

C 根据随机效应模型有效构造的统计量 W 服从自由度为 k-1 的有限卡方分布。即 $\text{var}(b-b) = \text{var}(b) - \text{var}(b) = W$

四、处理异方差问题

实际上，在处理面板数据线性回归时，主要考虑固定效应模型与 pooled OLS 的异方差问题。因为随机效应模型使用 GLS 估计，本身就已经控制了异方差。

GLS(广义最小二乘法)是一种常见的消除异方差的方法。它的主要思想是为解释变量加上一个权重，从而使得加上权重后的回归方程方差是相同的。因此在 GLS 方法下我们可以得到估计量的无偏和一致估计，并可以对其进行 OLS 下的 t 检验和 F 检验。