

固定效应: 目前看到解释的最清楚的帖子, 救命!

Fixed effects.



凡是搞计量经济的, 都关注这个号了

稿件: econometrics666@126.com

所有计量经济圈方法论丛的 code 程序, 宏微观数据库和各种软件都放在社群里. 欢迎到计量经济圈社群交流访问.



计量社群里在讨论固定效应相关问题时, 一位群友分享了一则题为《固定效应》的帖子。看罢, 群友们认为这是讲解最清楚易懂的干货帖, 值得阅读。

基于此, 遂将其分享给各位中青年学者参考, 来源附在文后。

关于固定效应, 参看: 1. [交互项! 交互项! 固定效应回归模型中的交互项!](#) 2. [在 Stata 中如何做 2SLS, DID, DEA, SFA, 面板 PSM, 二值选择, 固定效应和时间序列?](#) 3. [一定要控制时间固定效应吗?](#) 4. [公司和个体固定效应总是更好吗? 关于固定效应使用和解释的最全指南!](#) 5. [使用固定效应 FE 时良好做法对应的检查清单](#) , 6. [双向固定效应多期 DID 最新进展和代码汇总, 关于控制变量和固定效应选取的讨论](#) , 7. [快速估计带有高维固定效应的泊松模型, 这计算速度真快, 真实用!](#) 8. [不能直接控制某个固定效应时, 我们能尽量做些什么呢?](#) 9. [时间固定效应和时间趋势项的区别, 可以同时加?](#) 10. [省份 / 行业固定效应与年份固定效应的交乘项固定效应](#) , 11. [截面 DID, 各种固定效应, 安慰剂检验, 置换检验, 其他外部冲击的处理](#) , 12. [广义合成控制法 gsynth, 基于交互固定效应的因果推断](#) , 13. [固定效应模型 + 测量误差 = 有问题, 如何解决这问题呢?](#)

1. 固定效应是什么？

首先，固定效应是一种控制变量，可以影响解释变量，也可以影响被解释变量。其次，固定效应是有固定不变特征的影响因素（控制变量），包括可观测的和不可观测的。最后，固定效应有两个维度，一个是空间，一个是时间。

数据：首先是一个固定效应就是一个控制变量或者是一种综合的影响因素（包括了很多控制变量（产生的效果）），其次根据独一无二的固定不变的特征来分组，再次每组组内的取值相同，最后产生对组内个体同质性的影响。固定效应有几个类型：个体固定效应、时间固定效应、省份固定效应、行业固定效应以及交叉固定效应等等，这是因为根据独一无二的固定不变的特征来分组造成的，比如个体固定效应就是每个个体不随时间变化的独特特征对被解释变量的影响，省份固定效应就是每个省份不随时间变化的独特特征对被解释变量的影响，不同的分组可以形成不同的固定效应。

模型： $Y_{it}=bX_{it}+d_i+u_{it}$ ，个体固定效应就是 d_i ，类似于每个个体有一个单独的截距项，或者类似于常数项但是不同组有不同的取值，组内取值相同。或者 $Y_{it}=bX_{it} + \sum d_i D_i + u_{it}$ ，个体固定效应 d_i 就是虚拟变量 D_i 的系数。

图形：每个组分别画图，放在一个图内，不同的组有不同的截距项，但是所有组有相同的斜率。

2. 为什么要引入固定效应？

因为固定效应是一种控制变量，控制变量的目的就是防止遗漏变量带来的内生性问题，遗漏变量会影响解释变量，也会影响被解释变量，导致解释变量的估计系数不准确或者不纯，所以要多增加控制变量。控制变量一般有可观测的随时间变化的变量、有可观测的不随时间变化的变量，有不可观测不随时间变化的变量等等，固定效应主要控制不可观测不随时间变化的变量，比如个体固定效应，两个公司 A 和 B 绩效有大的不同，那么其背后有很多原因，比如公司文化、所处省份、所处行业等等，公司文化难衡量且不随时间变化，所以可以用个体固定效应来控制。当然也不一定是不可随时间变化，比如年份固定效应就是衡量每个年份不随个体变化的影响因素，每年都有独特的特征，如宏观经济波动导致企业 A 在不同年份的绩效不一样。

3. 举例

个体固定效应：个体固定效应就是个体层面不随时间变化的影响因素，每个个体都独特的个体特征，其目的是为了控制个体层面独一无二的不可随时间变化的特征，比如企业 A 之所以绩效比企业 B 好，其中就有企业文化在起作用，若是想要控制，就需要加入个体固定效应。写成模型就是 $Y_{it}=b*X_{it}+d_i+u_{it}$ ，这里的 d_i 就是个体固定效应，每个个体都有独特的部分，每个个体都不一样，不可随时间变化。

时间固定效应：时间固定效应就是时间层面不随个体变化的影响因素，每个年份都有独特的年份特征，其目的是为了控制时间层面独一无二的不可随个体变化的特征，比如每年的宏观经济。写成模型就是 $Y_{it}=b*X_{it}+T_t+u_{it}$ ，这里的 T_t 就是年份固定效应，每个年份都有独特的部分，每个年份都不一样，不可随个体变化。

一个具体的例子：我们想要研究税法改革与企业绩效之间的关系，企业 1 在 A 省 2014 年后发生了税法改革，企业 2 在 B 省一直都没有税法改革。我们如果仅仅看税法改革后企业 1 的绩效大于企业 2 的绩效，就说税法改革是有效的，显然不合理，其忽略了影响企业绩效的其他因素。比如企业 1 和企业 2 本身都有独特的个体固定特征，导致企业 1 本身就绩效高，这时候就需要控制个体固定效应，将原本就有

的每个企业的特色或者两企业之间的差异给估计出来控制住，这样才能进一步比较，这就是个体固定效应。又比如在 2014 年宏观经济因素发生了变化，但是与税法改革无关，导致企业 1 和企业 2 绩效都提升，那么还需要剥离宏观经济的影响，这时候就需要控制年份固定效应，将每年的特色给控制住。所以，2014 年后 A 企业平均绩效 = 税改 + 2014 年后宏观经济 + A 个体本身特色，2014 年后 B 企业平均绩效 = 2014 年后宏观经济 + B 个体本身特色，那么若想知道 2014 年后两个企业绩效的差距是否来自于税改，那么就需要控制住年份固定效应和个体固定效应。另外，前面企业个体固定效应是控制了企业本身独特的不随时间变化的特征，比如企业的文化、企业所在的省份和行业等等，企业所属的省份和行业也是影响企业绩效的关键因素，企业 A 和 B 之间绩效的不同也可能是因为省份的独特的不随时间变化的特征和行业的独特的不随时间变化的特征，按理说也应该一起控制，即省份固定效应和行业固定效应。那么到底要不要控制？等讲完固定效应的估计方法再说。

4. 怎么估计固定效应？

介绍两种方法：一种是虚拟变量法，另一种是组内离差法。

假设模型是 $Y_{it} = b \cdot X_{it} + d_i + u_{it}$ ， i 是个体如企业， t 是年份，这里的 d_i 就是个体固定效应——企业独一无二的、不随时间变化的影响解释变量和被解释变量的因素。可以说，我们要么就把 d_i 个体固定效应给估计出来，要么就把 d_i 个体固定效应给去掉后再估计。

4.1 虚拟变量法——直接估计出来

个体固定效应是每个企业独一无二的、不随时间变化的影响因素，那么可以放入虚拟变量来估计，在没有常数项时，一个企业就放入一个虚拟变量，如果个体是 i ，那么虚拟变量 $D_i = 1$ ，如果是其他个体，那么虚拟变量 $D_i = 0$ 。通过估计出虚拟变量的系数，就可以得到个体固定效应。

一个特点：当每个个体都有一个虚拟变量的时候，个体层面可以观测的、不随时间变化的虚拟变量，如性别变量，就会与个体固定效应虚拟变量共线性，这时候性别变量会 omitted 被剔除，个体固定效应实际上是可以观测和不可观测的、不随时间变化的影响因素。

代码：4 个企业个体

```
tab id, gen (id_dum) / 根据 id 生成虚拟变量
```

```
reg y x id_dum1-id_dum4,noconstant
```

4.2 组内离差法——先剔除后估计

我们知道每个个体都有不随时间变化的个体固定效应，一个个体有很多时间维度，就形成一个组，解释变量、被解释变量和控制变量对时间求均值，形成一个组内均值，但是个体固定效应由于不随时间变化，还是等于个体固定效应，那么变量的每个值都减去自己的组均值，就形成一个新变量，这时候新变量里就没有个体固定效应了。

原来模型： $Y_{it} = bX_{it} + d_i + u_{it}$ ，每组求均值后： $Y_i = bX_i + d_i + u_i$ ，相减后： $Y_{it} - Y_i = b(X_{it} - X_i) + u_{it} - u_i$ 。直接用 ols 就可以估计出来 b 。

一个特点：不仅去除了个体固定效应（不可观测的不随时间变化的变量），也去除了可观测的不随时间变化的变量，如性别变量，假设有两个个体，有 3 个年份，个体 1 是男，个体 2 是女，男取值为 1，女取值为 0，那么个体 1 求均值后为 1，个体 2 求均值后为 0，那么个体 1 的取值减去均值都变成了 0，个体 2 的取值减去均值都变成了 0，此时性别虚拟变量全部都成了 0，不存在个体内变化信息，无法估计。

代码：

```
egen m_y=mean(y),by(id) / 按 id 生成变量 y 均值
```

```
egen m_x=mean(x),by(id) / 按 id 生成变量 x 均值
```

```
gen winthin_y=y-m_y / 对 y 进行去均值转换
```

```
gen within_x=x-m_x/.....
```

```
gen winthin_y within_x,noconstant
```

以上是手动原理，以下是自动

```
xtset id year
```

```
xtreg y x, fe
```

5. 一些关于固定效应的问题

一种是固定效应之间的关系，另一种是固定效应与控制变量之间的关系。

5.1 有了个体固定效应还需不需要控制省份固定效应和行业固定效应？

看企业是否在年份变化里变动省份和行业。

重新假设一个数据，企业 1 和企业 2 在 A 省，企业 3 和企业 4 在 B 省，放入企业固定效应就是放入 4 个虚拟变量，放入省份固定效应就是放入 1 个虚拟变量，企业 1 和企业 2 的虚拟变量就可以表出省份虚拟变量，或者企业 3 和企业 4 的虚拟变量就可以表出省份虚拟变量，那么在放入回归的时候，省份虚拟变量会被 (omitted) 掉。或者按照组内离差的方法，要对每个个体变量（解释变量、被解释变量和控制变量）先根据个体取求均值，然后每个个体变量减去均值，然后跑回归。那么省份虚拟变量在取均值的时候还是等于一般的取值，所以一减就全是零了，不存在个体内变化信息，无法估计。那么什么时候可以同时放入省份和企业固定效应？企业虚拟变量表出不了省份虚拟变量的时候，也就是此时省份固定效应无法被代替了，此时个体变量减去组内均值还是存在的，可以回归。比如企业 1 在 2015 年之前在 A 省，2015 年之后在 B 省，这时候企业 1 的虚拟变量就没办法表出，控制的个体固定效应是个体不随时间变化的，省份变化了，所以就没控制。可以加入省份固定效应，因为之前个体固定效应所控制的省份固定效应不是真实的，或者就没控制省份固定效应，因为省份固定效应的控制需要整个省份整个年份的企业，但是由于企业的变化导致省份内企业数量的变化。

5.2 固定效应与控制变量之间的关系是什么？

其实无论个体固定效应、时间固定效应、省份、行业、省份 - 行业、省份 - 年份等等，都是将个体分成不同的组，固定效应就是要控制每个组内独一无二且固定不变的特征，当加入时间维度，就不一定是不随时间变化了，它变成了在那一个分组内固定不变，比如省份 - 年份的话，控制了浙江省 2012 年的独特的固定特征对。那么当加入个体固定效应时，个体层面的虚拟变量就会被剔除。当加入行业 - 年份固定效应时，行业层面随时间变化的控制变量就会被剔除。当加入省份 - 年份固定效应时，省份层面随时间变化的控制变量就会被剔除，如果你的核心解释变量是省份层面的，那么就不能加入省份 - 年份固定效应，因为解释变量会被剔除。

5.3 交互固定效应如何设定？

两个变量交互，比如行业 and 年份，文献中的表示方法一般是行业 - 年份或者行业 \times 年份。基本来说有两种方法：第一种方法先生成交乘变量，其实就是两个变量的进一步分组，比如行业与时间交互，有 3 个行业和 10 个年份，交互后就进一步分成 30 个组，同一个行业同一个年份的企业分成一组！也可以这么理解，将企业分到 3 个行业里面，一个行业又再按年份分成 10 组，那么就是 30 组。但是也不一定就是 $3 \times 3 = 9$ ，比如有三个行业和三个城市，先将企业分成 3 个城市，第一个城市有 3 个行业，但是第二个和第三个城市只有 1 个行业，且相同，那么所有的企业一共只能分成 5 组。那么代码是这样的：`egen year_industry = group(year industry)`，其次在 `reghdfe` 的 `absorb` 中写成 `absorb(year_industry)`；第二种方法就是直接在 `reghdfe` 的 `absorb` 中写成 `absorb(year#industry)`。其实本质上来说就是分组或者变换维度，形成一个新的类别变量或者维度，而且是更加细化的维度，那么最普遍的做法就是先形成一个新的变量即 `egen` 的做法，然后就把他当成普通的变量进行回归，如果要固定就固定，如果要聚类就聚类。

5.4 省份 - 年份固定效应与省份和年份固定效应之间的关系是什么？

其实，省份 - 年份就是用省份和年份一起来分组的，无论企业变动与否，都能完全分组，那么省份 - 年份可以以表出省份，也可以表出年份，所以控制了省份 - 年份就不用控制省份和年份了。在 `stata` 中，放入省份 - 年份虚拟变量后再放入年份虚拟变量会被剔除，放入省份虚拟变量也会被剔除。

一句话总结固定效应：有固定不变特征的控制变量，其目的是为了控制每个分组内独一无二的固定不变的特征。

6. 控制几个固定效应？

控制变量有两个维度，一个是时间，另一个是空间，空间又分省份、行业和企业个体。

在个体层面，分为随时间变化的控制变量和不随时间变化的控制变量，随时间变化的控制变量可以用可观测的控制变量来衡量，不随时间变化的控制变量可以用个体固定效应来衡量。那么个体层面控制变量 = 个体可观测控制变量 + 个体固定效应。

在省份层面，分随时间变化的控制变量和不随时间变化的控制变量，随时间变化的控制变量可以用可观测的控制变量来衡量，不随时间变化的控制变量可以用省份固定效应来衡量。那么省份层面控制变量 = 省份可观测控制变量 + 省份固定效应。

在行业层面，分随时间变化的控制变量和不随时间变化的控制变量，随时间变化的控制变量可以用可观测的控制变量来衡量，不随时间变化的控制变量可以用行业固定效应来衡量。那么行业层面控制变量 = 行业可观测控制变量 + 行业固定效应。

在时间层面，如果想要控制不随个体变化、不随省份变化、不随行业变化的控制变量，可以时间固定效应。

那么第一种加入控制变量的方式就是：个体层面控制变量 + 省份层面控制变量 + 行业层面控制变量 + 时间层面控制变量 = 个体可观测控制变量 + 省份可观测控制变量 + 行业可观测控制变量 + 个体固定效应 + 省份固定效应 + 行业固定效应 + 时间固定效应。

如果省份可观测控制变量不太够，那么可以换成省份 - 年份固定效应，此时第二种加入控制变量的方式就是：个体可观测控制变量 + 行业可观测控制变量 + 个体固定效应 + 行业固定效应 + 省份 - 年份固定效应。此时如果解释变量是省份层面的，那么不能加入省份 - 年份固定效应，因为解释变量会被剔除。

如果行业可观测控制变量不太够，那么可以换成行业 - 年份固定效应，此时第三种加入控制变量的方式就是：个体可观测控制变量 + 省份可观测控制变量 + 个体固定效应 + 省份固定效应 + 行业 - 年份固定效应。此时如果解释变量是行业层面的，那么不能加入行业 - 年份固定效应，因为解释变量会被剔除。

Source: https://zhuanlan.zhihu.com/p/476735732?utm_medium=social&utm_oi=1134048306945794048，作者：胜风。

1. [用 "因果关系图" 来进行因果推断的新技能](#)，2. [因果推断专题：因果图](#)，3. [因果推断专题：有向无环图 DAG](#)，4. [confounder 与 collider 啥区别？混淆 vs 对撞](#)，5. [三张图秒懂，混淆，中介，调节，对撞，暴露，结果和协变量的复杂关系](#)，6. [中介效应检验流程，示意图公布，不再畏惧中介分析](#)，7. [图灵奖得主 Pearl 的因果推断新科学，Book of Why?](#) 8. [前沿：nature 刊掀起 DAG 热，不掌握就遭淘汰无疑！因果关系研究的图形工具！](#) 9. [前沿：卫星数据在实证研究中的应用，用其开展因果推断的好处！](#) 10. [7 大因果推断大法精选实证论文，可用于中国本土博士课堂教学！](#) 11. [随机分配是什么，为什么重要，对因果关系影响几何？](#) 12. [应用计量经济学现状：因果推断与政策评估最全综述](#)，13. [疫情期计量课程免费开放！面板数据，因果推断，时间序列分析与 Stata 应用](#)，14. [Python 做因果推断的方法示例，解读与 code](#)，15. [内生转换模型 vs 内生处理模型 vs 样本选择模型 vs 工具变量 2SLS](#)，16. [不用 IV，基于异方差识别方法解决内生性，赐一文献](#) 等等。

下面这些短链接文章属于合集，可以收藏起来阅读，不然以后都找不到了。

4 年，计量经济圈近 1500 篇不重类计量文章，

[可直接在公众号菜单栏搜索任何计量相关问题](#)，

Econometrics Circle

数据系列：[空间矩阵](#) | [工企数据](#) | [PM2.5](#) | [市场化指数](#) | [CO2 数据](#) | [夜间灯光](#) | [官员发言](#) | [微观数据](#) | [内部数据](#)

计量系列：[匹配方法](#) | [内生性](#) | [工具变量](#) | [DID](#) | [面板数据](#) | [常用 TOOL](#) | [中介调节](#) | [时间序列](#) | [RDD 断点](#) | [合成控制](#) | [200 篇合集](#) | [因果识别](#) | [社会网络](#) | [空间 DID](#)

数据处理：[Stata](#) | [R](#) | [Python](#) | [缺失值](#) | [CHIP/CHNS/CHARLS/CFPS/CGSS](#) 等 |

干货系列：[能源环境](#) | [效率研究](#) | [空间计量](#) | [国际经贸](#) | [计量软件](#) | [商科研究](#) | [机器学习](#) | [SSCI](#) | [CSSCI](#) | [SSCI 查询](#) | [名家经验](#)

计量经济圈组织了一个计量社群，有如下特征：热情互助最多、前沿趋势最多、社科资料最多、社科数据最多、科研牛人最多、海外名校最多。因此，建议积极进取和有强烈研习激情的中青年学者到社群交流探讨，始终坚信优秀是通过感染优秀而互相成就彼此的。



全文完