



南京大學  
NANJING UNIVERSITY



武漢大學  
WUHAN UNIVERSITY

2021年Stata讲座：

战略管理研究中的PSM-DID设计与实现

刘海建 南京大学

## 一、倾向性匹配得分 ( Propensity Score Matching : PSM)

- 在经济学中，我们通常希望评估某项公共政策实施后的效应，为此，我们构建 "处理组" 和 "控制组" 以评估处理效应 ( **treatment effect** )。然而，我们的数据通常来自非随机的观察研究中，处理组和控制组的初始条件不完全相同，故存在选择偏差 ( **selection bias** ) 问题。
- 倾向得分匹配法( **PSM** )使用倾向得分函数将多维向量的信息压缩到一维，然后根据倾向得分进行匹配。这样可以在既定的可观测特征变量下，使得处理组个体和控制组个体尽可能相似，因而可以缓解处理效应的选择偏差问题。

## 1. PSM的模型设定

对于个体*i*，根据是否进行某项处理可以分为两种结果：

$$y_i = \begin{cases} y_{1i}, & \text{if } D_i = 1 \\ y_{0i}, & \text{if } D_i = 0 \end{cases} \quad (1)$$

$D_i$  表示个体  $i$  是否进行某项处理，即 **1** 表示处理，**0** 表示未处理；

$y_{1i}$  表示个体进行处理的结果；

$y_{0i}$  表示个体未进行处理的结果。

在给定可观测特征变量  $x_i$  情况下，个体  $i$  进入处理组的条件概率为：

$$p(x_i) = \Pr(D_i = 1 \mid x = x_i) = E(D_i \mid x_i) \quad (2)$$

根据式 (1) 和 (2) 可得参与者的平均处理效应 (ATT) 为：

$$\begin{aligned} ATT &\equiv E\{y_{1i} - y_{0i} \mid D_i = 1\} \\ &= E[E\{y_{1i} - y_{0i} \mid D_i = 1, p(x_i)\}] \\ &= E[E\{y_{1i} \mid D_i = 1, p(x_i)\} - E\{y_{0i} \mid D_i = 0, p(x_i)\} \mid D_i = 1] \end{aligned} \quad (3)$$

## 2. PSM的假设条件

### ➤ 共同支撑假设 ( Common Support Assumption )

对于 $x$ 的任何可能取值, 都有  $0 < p(x) < 1$ 。此假定意味着处理组和控制组这两个子样本存在重叠, 这是进行匹配的前提。它保证了处理组与控制组的倾向得分取值有相同的部分。

### ➤ 平行假设 ( Balancing Assumption )

给定 $p(x_i)$ , 则  $D_i$ 独立于  $x_i$ 。此假定意味着, 对于给定的  $p(x_i)$ , 处理是随机的, 即在接受处理之前, 处理组和控制组之间没有差异, 处理组产生的效应完全来自处理。

### 3. PSM 一般步骤

- (1) 选择协变量  $x_i$  : 尽可能涵盖影响  $(y_{0i}, y_{1i})$  与  $D_i$  的相关变量;
- (2) 估计倾向得分值: 可以使用 **probit** 或 **logit** 模型估计;
- (3) 检验平行假设是否满足: 使得  $x_i$  在匹配后的处理组均值和控制组均值较接近, 保证数据平衡;
- (4) 根据倾向得分值将处理组个体和控制组个体进行配对: 匹配的方法有最近邻匹配、半径匹配、核匹配等;
- (5) 根据匹配后样本计算 ATT。

## 4. PSM的局限性

PSM试图通过匹配再抽样的方法使得观测数据尽可能地接近随机实验数据，在很大程度上减少了观测数据的偏差，但它本身也有如下的局限性：

- (1) PSM通常要求比较大的样本容量以得到高质量的匹配。
- (2) PSM要求处理组与控制组的倾向得分有较大的共同取值范围 ( **common support** ); 否则，将丢失较多观测值，导致剩下的样本不具有代表性。
- (3) PSM只控制了可测变量的影响，如果存在依不可测变量选择 ( **selection on unobservable** ) ，仍会带来“隐性偏差” ( **hidden bias** )。

## 二、双重差分模型 (Differences-in-Differences, 简称 DID)

### 1. DID 简介

- DID 作为计量经济学的一种重要方法, 被广泛应用于各项经济政策、卫生政策等效果评估之中, 并取得了较好效果 (Bertrand, Duflo & Mullainathan, 2004; Lechner, 2011; Dimick & Ryan, 2014)
- 含义: 为估计处理效果, 比较处理后与处理前的差异, 利用处理组的前后变化减去控制组的前后变化, 就是处理效应, 故名"双重差分"。
- 优点: (1) 可以很大程度上避免内生性问题的困扰: 政策相对于微观经济主体而言一般是外生的, 因而不存在逆向因果问题。此外, 使用固定效应估计一定程度上也缓解了遗漏变量偏误问题。

(2) 传统方法下评估政策效应, 主要是通过设置一个政策发生与否的虚拟变量然后进行回归, 相较而言, 双重差分法的模型设置更加科学, 能更加准确地估计出政策效应。

## 2. 双重差分法的基本思想：

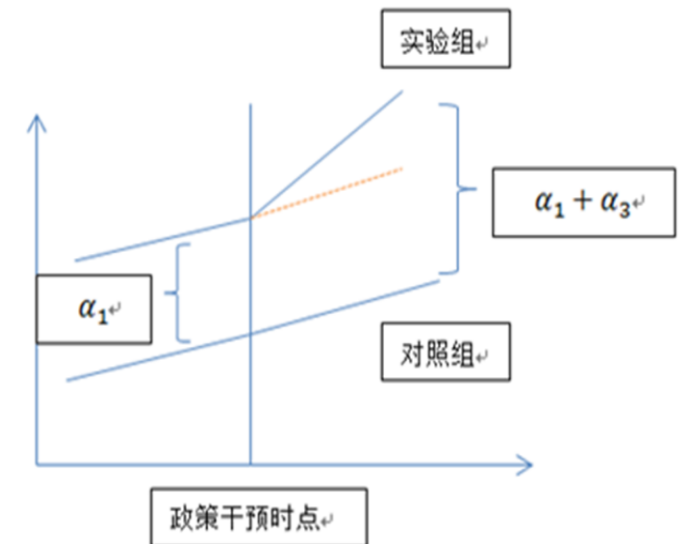
- 使用“未受干预个体”来估计受干预个体在不受干预的情况下潜在结果的变化-- 其他因素的影响；
- 从“受干预个体”干预前后结果变化中剔除其他因素的影响，便可以得到干预效应。

## 3. 双重差分法的基本假设：

共同趋势假设 (Common Trends Assumption) :

干预组个体在没有收到干预的情况下的变动趋势与未受干预组相同。

- 时间对干预组和控制组潜在结果的效应是“相加可分离 (additive and separable)”的，
- 时间对干预组和控制组潜在结果的效应是相同的。



图中橙色虚线表示的是假设政策并未实施时处理组的发展趋势。事实上，该图也反映出了DID最为重要和关键的前提条件：共同趋势 (Common Trends)，也就是说，处理组和对照组在政策实施之前必须具有相同的发展趋势。



## 4.传统DID模型

DID的使用场景为，在面板数据中，个体可分为两类，即受到政策冲击的“处理组”（treatment group）与未受政策影响的“控制组”（control group）。重点落在政策冲击和是否受到政策冲击，通过引入虚拟变量来实现。即：政策冲击前后（pre-post）设为0和1，是否受到政策冲击（control-treat）设为0和1。

$$\begin{aligned}
 treat_i &= \begin{cases} 1 & \text{if } i \in \text{treatment group} \\ 0 & \text{if } i \in \text{control group} \end{cases} & post_t &= \begin{cases} 1 & \text{if } t \in \text{post treatment} \\ 0 & \text{if } t \in \text{before treatment} \end{cases}
 \end{aligned}$$

因而，模型常设计为：

$$y_{it} = \alpha + \beta treat_i + \gamma post_t + \delta treat_i \times post_t + \boldsymbol{\eta}' \mathbf{x}_{it} + \varepsilon_{it}$$

处理组虚拟变量 $\beta$ 捕捉了处理组的组别效应（处理组与控制组的固有差别），处理期虚拟变量 $\gamma$ 控制了处理期的时间效应（处理期前后的固有时间趋势）， $\mathbf{x}$ 为其他控制变量，而交互项 $\delta$ 则代表了处理组在处理期的真正效应（受到政策冲击的效应），这正是我们关心的处理效应。

## 5.经典DID模型

经典DID是在传统DID模型上控制了个体固定效应 (individual fixed effects) 和时间固定效应 (time fixed effects) , 并去除单独变量。模型如下:

$$y_{i,t} = \alpha + \mu_i + \lambda_t + \theta treat_i \times post_t + \beta x_{i,t} + \epsilon_{i,t}$$

(1)  $\mu$ 为个体固定效应。加入个体固定效应之后, 就不必再放入处理组虚拟变量 ( $treat_i$ ) , 否则会引起多重共线性问题。因为前者包含比后者更多的信息 (前者控制到个体层面, 而后者仅控制到组别层面) 。

(2)  $\lambda$ 为时间固定效应。同理, 加入时间固定效应就不用再加处理期虚拟变量 ( $post_t$ ) 。否则, 将导致严格多重共线性, 因为前者包含比后者更多的信息 (前者控制了每一期的时间效应, 而后者仅控制处理期前后的时间效应)

(3) 交乘项的系数 $\theta$ 是我们关心的处理效应。系数 $\theta$ 为:  
上式的含义是处理组前后变化与控制组前后变化之差,  
就是处理效应。

$$\begin{aligned} \theta &= \{E[y_{i,t} | treat_i = 1, post_t = 1] - E[y_{i,t} | treat_i = 1, post_t = 0]\} \\ &\quad - \{E[y_{i,t} | treat_i = 0, post_t = 1] - E[y_{i,t} | treat_i = 0, post_t = 0]\} \\ &= (Y_1 - Y_0) - (C_1 - C_0) = (\theta + \lambda_t) - \lambda_t \\ &= (Y_1 - C_1) - (Y_0 - C_0) = (\theta + \mu_i) - \mu_i \end{aligned}$$

## 6.DID模型的拓展

当前，越来越多的学者开始投入对DID设计的理论研究，这一研究围绕着技术路径和政策路径两方面展开。

- 技术路径是探讨放宽DID经典假设的解决策略，如当平行趋势假设不能够满足时，如何对干预效果进行估计（Athey & Imbens, 2018; Strezhnev, 2018）。
- 政策路径是探究新的DID设计，以解决政策评估中新问题，最典型的是如何处理政策干预时间差异的因果效应问题（Callaway & Sant'Anna, 2018）。

## 7.DID模型局限

- (1) 对数据的要求，可以获得政策执行前的数据以及政策执行后的数据；
- (2) 要求可以找到好的对照组（即除了是否实施该政策有有差异外，其余的个体差异应当控制住并保持不变）

### 三、PSM+DID

近年来，经济学、管理学研究出现了“自然实验化”的趋势，相关计量经济学模型也得到了快速发展。其中，倾向得分匹配-双重差分模型 (以下简称 **PSM-DID**) 作为有力的政策分析工具更是被广泛使用。

**PSM-DID** 模型是由倾向得分匹配模型 (**PSM**) 和双重差分模型 (**DID**) 结合而成。其中，**PSM** 模型根据倾向得分寻找与处理组尽可能相似的控制组，**DID** 模型负责识别政策冲击所产生的影响。

### 三、PSM+DID

- 与PSM相比，PSM-DID模型除了可以控制可测变量产生的个体差异，还可以控制不可测因素产生的个体差异；
- 与DID相比，PSM-DID模型考虑到了实验组和控制组的个体差异；

因此，PSM-DID模型主要解决的问题是：

- (1) 研究目的是某项政策或策略实施的效果，有时间前后的对比；
- (2) 实验组和控制组存在个体差异问题，不满足DID的平行趋势假定；
- (3) 可以观测到政策执行前后以及政策执行与否的结果变量；
- (4) 受到不可观测变量的影响，存在选择性偏误或者遗漏变量问题。

## 四、具体实例

### 高管团队成员死亡与善待员工--一个死亡反思的视角

- 研究问题：高管死亡这类突发事件会如何影响善待企业内部员工？
- 方法：PSM+DID
- 样本选择：实验组：2010-2018年我国396名上市企业高管团队成员死亡的数据；  
 控制组：参考陈国立等（2019）为经历过董事死亡事件的公司寻找对照组的方法，选取董事会规模、董事平均年龄、高管团队平均年龄、企业规模和总资产净利润率五个变量为匹配变量来选择控制组。
- 模型：（1）PSM的logit模型：

$$\text{Treat}_{i,t} = \beta_0 + \beta_1 \text{director\_number}_{i,t} + \beta_2 \text{director\_age}_{i,t} + \beta_3 \text{Eexecutive\_age\_ownership}_{i,t} \\ + \beta_4 \text{Firm\_age}_{i,t} + \beta_5 \text{ROA}_{i,t} + \text{Industry} + \text{Year} + \varepsilon_{i,t}$$

## 四、具体实例

### 高管团队成员死亡与善待员工--一个死亡反思的视角

(2) DID模型:

$$ET = \alpha + \beta_1 \text{treated} * \text{posted} + \beta_2 \text{controls} + \mu_i + \lambda_t + \varepsilon_{i,t}$$

其中ET为衡量善待员工的指标；treated为实验组虚拟变量，实验组取值为1，否则为0；posted代表事件发生的虚拟变量，当年度处于有高管团队成员死亡之后取值为1，否则为0； $\beta_1$ 刻画了高管团队成员死亡事件对员工待遇的影响，是本研究主要关注的对象。Controls为一组控制变量； $\mu$ 表示个体固定效应； $\lambda$ 表示时间固定效应。

## 四、具体实例

### 高管团队成员死亡与善待员工--一个死亡反思的视角

(3) 调节效应模型:

调节变量: 高管平均年龄; 死亡事件性质

$$ET = \alpha + \beta_1 \text{treated} * \text{posted} * \text{executive\_age} + \beta_2 \text{controls} + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (2)$$

用模型 (2) 来检验高管平均年龄是否影响高管团队成员死亡事件与员工待遇之间的关系。

$$ET = \alpha + \beta_1 \text{treated} * \text{posted} * \text{sudden\_died} + \beta_2 \text{controls} + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (3)$$

用模型 (3) 来检验死亡事件性质为猝死时是否会影响高管团队成员死亡事件与员工待遇之间的关系。



***Thank You!***

