THE WORLD BANK

IMPACT EVALUATION

技术交流会（六）
匹配法

中国·北京

2009

Human Development Network

Spanish Impact Evaluation Fund

什么情况下应用匹配法?

- 如果不是随机试验，而是观察性研究时怎么办？？
  - *这种情况下就可以应用匹配法！如果试验对象是建立在可观测变量的基础上，则可以在此基础上建立对照组。*

- 注意：如果是不可观测的，则匹配法**仍然**不能用来控制选择偏倚。

- 直观：试验之前，应根据观察使对照组与干预组尽可能相似。
- 该方法假定在干预组和对照组之间，已经不存在不可观测的差异。

关键问题....

当试验对象是可观测变量时，干预措施对干预组会有哪些影响？

无混杂和观测值选择

- 假设*X*表示一个矩阵，矩阵中每行都表示个体i预处理可观测变量的一个向量。
- 无混杂的定义

  如果预处理可观测变量X满足：

  $$Y_1, Y_0 \perp D \mid X$$

  则试验对象是无混杂的。

- 也就是说，无混杂满足以下条件
  - 对*X* 中的每一个体：随机处理
  - 入选条件仅依赖于对X的观测值

如果没有混杂，干预措施对干预组的平均效果

- 直观:
  - 估计干预措施对由X定义的每一对象的效果；
  - 取所有单元效果估计值的均值。

- 数学证明：见附件1。

估计平均效果的方法-可观测变量的选择

- 无混杂性意味着可以用以下方法来估算干预的平均效果δ
  - 将数据分层选入由X的每一特征值确定的单元中
  - 对每个单元，计算其干预组和对照组干预效果均值的差异
  - 根据X 在干预组中的分布，计算以上差值的均值。
- 这种方法是否可行?

我们的方法是否可行？ …. 维度问题

- 当以下情况存在时，可能不可行…
  - 样本量太小，
  - 协变量范围太大，
  - 很多协变量有多个值，或者是连续的。
- 这就是所谓的维度问题…

维度问题

- 例如:
  - 2 个 二分类 变量X有多少单元? 3 个 二分类X 呢? K个 二分类X 呢?
  - 如果我们有两个变量，每个变量都有7个值呢?
- 随着单元数的增多，我们将“缺少共同的支集”
  - 仅包含干预组单元
  - 仅包含对照组单元

解决维度问题的一种选择

- 在*倾向评分*概念基础上，Rosenbaum and Rubin (1983) 提出一种可行的类似估算方法
  - 倾向评分可以将多维问题转化为单维问题。
  - 这样就减少了维度问题。

基于倾向评分的匹配

- 倾向评分**定义**:*倾向评分*是预处理变量接受干预的条件概率。

$$p(X) = Pr\{D = 1|X\} = E_X\{D|X\}$$

- **引理 1: 如果** $p(X)$ **是倾向评分,** ***那么*** $D \perp X \mid p(X)$

“在倾向性评分中，预处理变量会在有效与无效 之间获得平衡”

- **引理 2:** $Y_1, Y_0 \perp D \mid X \quad => \quad Y_1, Y_0 \perp D \mid p(X)$

“如果对预处理变量X的干预是非混杂的，那么倾向评分p(X)也是非混杂的。”

倾向评分能够解决维度问题吗?

- 能！
- 倾向评分的*平衡特性*（引理1）保证了：
  - 具有相同倾向评分观察变量的可观测协变量是同分布的，不受试验条件影响；
  - 对特定的倾向评分：试验处理是“随机的”，因此在观察中，对干预组与对照组是均衡的。

估算方法的运用

干预效果平均值δ的计算方法如下：

- 步骤 1: 估算倾向评分值
  - 例如利用logit函数，见附件3
  - 这一步骤是很必要的，因为倾向评分的“真实”值是未知的，因此必须要估算倾向评分值。
- 步骤 2: 在倾向评分基础上，估算干预效果的平均值。

什么情况下应用倾向评分匹配?

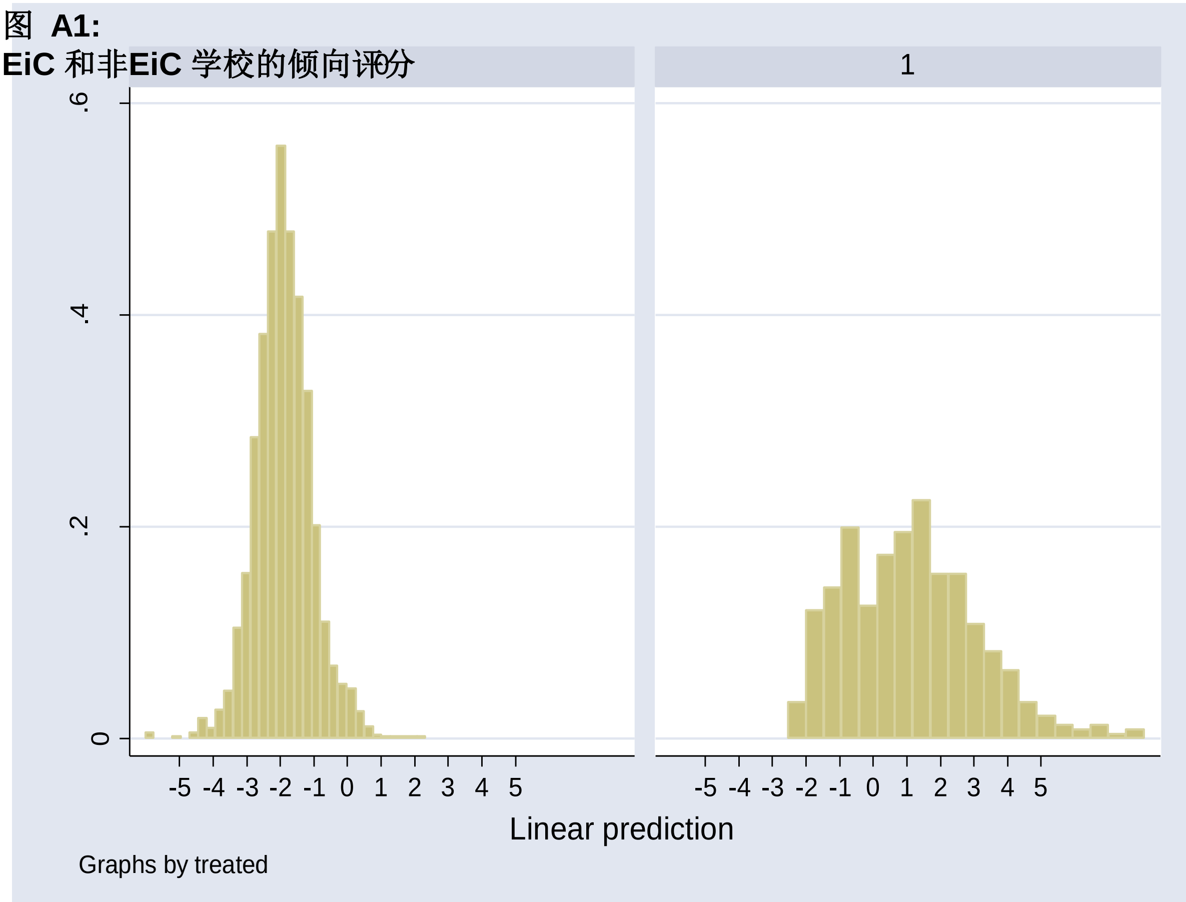
- 倾向评分的背后思想：估测干预效果需要对干预组和对照组进行严格匹配。
- 如果观察中干预组和对照组的差异很大，那么这种匹配是不充分的或不可信的，甚至是不可能的。
- 干预组和对照组倾向评分值的对比，为评价干预组和对照组的相似性提供了依据，因此也可以推断估算方法的可信度。

因此，您希望干预组和对照组的倾向评分是“相同的”…

- 干预组和对照组倾向评分值的变化范围应当是一致的。
  - 计算一下，对照组有多少对象的倾向评分值低于干预组的最小值，或者有多少高于干预组的最大值。
  - 反之亦然
- 干预组和对照组倾向评分值的频数分布是一致的。
  - 绘制估算出的干预组与对照组的倾向评分值直方图。
  - 直方图内的各个对应值形成了倾向评分的估算值。

共同的支集的实例：

图 **A1:**
EiC 和非**EiC** 学校的倾向评分



Source Machin, McNally, Meghir,
“卓越的城市教育” **EXCELLENCE IN CITIES:** 贫困地区一种教育政策的评估。

估算方法的实施

我们正在讨论估算干预效果δ的方法

- **步骤 1: 估算倾向评分(见附件3)**
- **步骤2: 在倾向评分基础上，估算干预效果的均值。**
  - 对（估算的）倾向评分值相同的干预组和对照组进行配对；
  - 根据每组（估算的）倾向评分值，计算干预效果；
  - 计算干预效果的平均值。

步骤2: 在倾向评分基础上，估算干预效果的均值

- 这在实践中是不可行的，因为很难找到两个倾向评分相同的单元。
- 我们所能做的最佳尝试是：将倾向评分*最接近的*干预组和对照组进行匹配。
- “最接近”可以有多种定义。不同的定义对应不同的匹配方法….
  - 根据评分分层;
  - 评分最接近的相匹配;
  - 评分半径匹配;
  - 评分的Kernel匹配;
  - 评分基础上的权重赋值。

参考文献：

- Dehejia, R.H. and S. Wahba (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94, 448, 1053-1062.
- Dehejia, R.H. and S. Wahba (1996), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", Harvard University, Mimeo.
- Hahn, Jinyong (1998), " ON the role of the propensity score in efficient semiparamentric estimation of average treatment effects", *Econometrica*, 66,2,315-331.
- Heckman, James J. H. Ichimura, and P. Todd (1998), " Matching as an econometric evaluation estimator ", *Review of Economic Studies*, 65, 261-294.
- Hirano, K., G.W. Imbens and G. Ridder (2000), "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score", mimeo.
- Rosenbaum, P.R. and D.B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika* 70, 1, 41–55.
- Vinha, K. (2006) A primer on Propensity Score Matching Estimators" Documento CEDE 2006-13, Universidad de los Andes

附件 1: 假设不存在混杂的情况下，对干预组干预效果的平均值

如果我们假设不存在混杂因素：

$$E_i\left\{Y_0\left(u_i\right)|\,D_i=0,X\right\}=E_i\left\{Y_0\left(u_i\right)|\,D_i=1,X\right\}=E_i\left\{Y_0\left(u_i\right)|\,X\right\}$$

$$E_i\left\{Y_1\left(u_i\right)|\,D_i=0,X\right\}=E_i\left\{Y_1\left(u_i\right)|\,D_i=1,X\right\}=E_i\left\{Y_1\left(u_i\right)|\,X\right\}$$

利用这种方式，我们可以为每个X单元定义：

$\delta_X =$ X单元的干预平均效果

$$= E_i\{\Delta_i\,|\,D_i=1,X\}$$

$$= E_i\left\{Y_1\left(u_i\right)-Y_0\left(u_i\right)|\,D_i=1,X\right\}$$

$$= \underbrace{E_i\left\{Y_1\left(u_i\right)|\,D_i=1,X\right\}}_{\text{can measure sample analog}}-\underbrace{E\left\{Y_0\left(u_i\right)|\,D_i=1,X\right\}}_{\text{can NOT measure sample analog}}$$

$$= E_i\left\{Y_1\left(u_i\right)|\,D_i=1,X\right\}-\underbrace{E_i\left\{Y_0\left(u_i\right)|\,D_i=0,X\right\}}_{\text{can measure sample analog}}$$

附件 1: 假设不存在混杂的情况下，对干预组干预效果的平均值

二者之间有何联系？

δ “干预组的干预效果平均值” …和…

δ_X “对干预组中X单元的干预效果平均值”

δ = 干预组干预效果平均值

$$= E_i\{\Delta_i \mid D_i = 1\}$$

$\Downarrow$ 根据迭代期望法则

$$= E_i\{E_X[\Delta_i \mid D_i = 1, X]\}$$

$$= E_X\{E_i[\Delta_i \mid D_i = 1, X]\}$$

$$= E_X\{\delta_X\}$$

$= E_X$ {干预组中X单元的平均影响}

附件2: 干预效果平均值和倾向评分

在使用倾向评分p(X)而非X矩阵的基础上，
我们先对干预组和对照组进行匹配

$$E_i\left\{Y_0\left(u_i\right) \mid D_i=0, p\left(X_i\right)\right\}=E_i\left\{Y_0\left(u_i\right) \mid D_i=1, p\left(X_i\right)\right\}=E_i\left\{Y_0\left(u_i\right) \mid p\left(X_i\right)\right\}$$

$$E_i\left\{Y_1\left(u_i\right) \mid D_i=0, p\left(X_i\right)\right\}=E_i\left\{Y_1\left(u_i\right) \mid D_i=1, p\left(X_i\right)\right\}=E_i\left\{Y_1\left(u_i\right) \mid p\left(X_i\right)\right\}$$

利用这些表达，我们通过p(X)可以定义f单元

$\delta_{p(X)}$ = p(X)定义单元的干预效果均值

$$\begin{aligned}
&= E_i\{\Delta_i \mid D_i=1, p(X)\} \\
&= E_i\left\{Y_1\left(u_i\right)-Y_0\left(u_i\right) \mid D_i=1, p(X)\right\} \\
&= \underbrace{E_i\left\{Y_1\left(u_i\right) \mid D_i=1, p(X)\right\}}_{\text{can measure sample analog}}-\underbrace{E\left\{Y_0\left(u_i\right) \mid D_i=1, p(X)\right\}}_{\text{can NOT measure sample analog}} \\
&= E_i\left\{Y_1\left(u_i\right) \mid D_i=1, p(X)\right\}-\underbrace{E_i\left\{Y_0\left(u_i\right) \mid D_i=0, p(X)\right\}}_{\text{can measure sample analog}}
\end{aligned}$$

附件2: 干预效果平均值和倾向评分

二者之间有何联系？

δ “干预组的干预效果平均值” …和…

δ_X “对干预组中X单元的干预效果平均值”

δ = 干预组干预效果平均值

$= E_i\left\{\Delta_i \mid D_i = 1\right\}$

$\Downarrow$ 根据迭代期望法则

$= E_i\left\{E_X\left[\Delta_i \mid D_i = 1, X\right]\right\}$

$= E_X\left\{E_i\left[\Delta_i \mid D_i = 1, X\right]\right\}$

$= E_X\left\{\delta_X\right\}$

$= E_X$ {干预组中X单元的平均影响}

附件 3: 倾向评分的估算

- 任何标准概率模型都可以用来估算倾向评分值，例如logit模型：

$$Pr\{D_i \mid X_i\} = \frac{e^{\lambda h(X_i)}}{1+e^{\lambda h(X_i)}} \qquad (16)$$

公式中$h(X_i)$ 为一个线性或者更高阶项协变量的函数。

倾向评分值的估算

- $h(X_i)$中包含哪些更高阶项?
  - 这由获取倾向评分值估算的需要决定，这种估算应当满足*平衡特性*。
- $h(X_i)$ 的规范：
  - 比X观察量之间的全部交互作用更小
  - 如果交互作用不很小：它仍需要满足平衡特性。
- 注意：倾向评分值的估计不需要行为解释

一种估算倾向评分值的运算法则

1. 从一个精简的logit 或者probit概率函数开始，测算评分值。
2. 根据倾向评分值，将数据排序（由低到高）。
3. 对模块中的所有观察值进行分层，使每个模块中的干预组和对照组的倾向评分值都无统计差异：
   a. 按评分值划分为5个等距模块{0 - 0.2, ..., 0.8 - 1}
   b. 对每个模块，检验干预组和对照组评分的均值是否 有统计学差异
   c. 如果有，增加模块数量然后继续检验
   d. 如果没有，进行下一步操作。

一种估算倾向评分值的运算法则（续）

4. 检验平衡特性对所有协变量在所有模块中的约束作用:
   a) 对每个协变量，检验所有模块中干预组和对照组均值（或更高阶项）是否有统计学差异；
   b) 如果一个协变量在某个模块中不平衡，那么将这个模块拆分，然后继续在更好的模块之间进行比较；
   c) 如果有一个协变量在所有模块中都不平衡，那么修改倾向评分的logit估计，增加更多干预和高阶项，再检验。

注意在所有这些程序中，结果都无作用

应用STATA程序pscore.ado

下载网址：http://www.iue.it/Personal/Ichino/Welcome.html