

倾向得分匹配法

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。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，故存在选择偏差问题。倾向得分匹配 (PSM) 法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。该方法最早由统计学家 Rosenbaum and Rubin(1983) 提出。

倾向得分计算

一般使用logit回归估计 $\hat{p}(x)$ 且估计包含 x 的高次方项

$$P(D_i | X = x) = E(D_i) = \frac{e^{x_i\beta_i}}{1 + e^{x_i\beta_i}} = \frac{1}{1 + e^{-x_i\beta_i}} \quad (1)$$

$$\log\left(\frac{p}{1-p}\right) = x_i\beta_i \quad (2)$$

设给定协变量 (x_0, x_1) :

$$p = \frac{e^{x_i\beta_i}}{1 + e^{x_i\beta_i}} \quad (3)$$

$$1 - p = \frac{1}{1 + e^{x_i\beta_i}} \quad (4)$$

$$\log L(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^N W_i(\beta_0 + \beta_1 + \beta_2) - \sum_{i=1}^N \log(1 + \exp(\beta_0 + \beta_1 + \beta_2)) \quad (5)$$

上式对 β 的偏导数使得该似然函数最大化。实际研究中,这一问题很少通过解析方式来解决,而是经常依赖于数值程序 (numerical procedure)来找到 β 的估计值。常见的3种数值估计法分比为:Newton-Raphson法、赋值法 (the scoring method)以及B-triple-H法(BHHH)。通常情况下，一个数值法 (numerical method)涉及以下步骤:

- 在式3右边中代入 β_0 、 β_1 和 β_2 的初始值(starting value)(即主观猜测值(guesses))以得到式3的第一个猜测值;
- 在式3右边中代入一组不同的 β_0 、 β_1 和 β_2 以获得上式的第二个猜测值;通过比较将新的log L和旧的 log L，这样我们可以知道尝试下一组 β_0 、 β_1 和 β_2 的方向;从步骤a到步骤b的过程被称作一次迭代(iteration);
- 重复上述过程若干次(即进行若干次迭代),直到得到log L的最大值(即最大的对数似然函数)或者直到两次迭代之间log L上的差值不再大于某一事先设定的标准值，比如 $\epsilon = 0.000001$ 。

倾向得分匹配法的分析过程

倾向得分匹配模型设定

对于个体 i ，根据是否进行某项处理可以分为两种结果：

$$y_i = \begin{cases} y_{1i}, & \text{if } D_i = 1 \\ y_{0i}, & \text{if } D_i = 0 \end{cases} \quad (6)$$

D_i 表示个体 i 是否进入某项处理，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。

定义个体 i 的倾向得分为：在给定的 x_i 的情况下，个体 i 进入处理组的条件概率，即：

$$p(x_i) = \Pr(D_i = 1 | x = x_i) = E(D_i | x_i) \quad (7)$$

根据式 (6) 和式 (7) 可得参与者的平均处理效应 (ATT) 为：

$$\begin{aligned} ATT &\equiv E\{y_{1i} - y_{0i} | D_i = 1\} \\ &= E[E\{y_{1i} - y_{0i} | D_i = 1, p(x_i)\}] \\ &= E[E\{y_{1i} | D_i = 1, p(x_i)\} - E\{y_{0i} | D_i = 0, p(x_i)\} | D_i = 1] \end{aligned} \quad (8)$$

反事实框架

由于数据的可获得性不同，当我们使用未被处理baseline时，模型中可能存在自选择偏差、遗漏变量偏差、反向因果造成的偏差。

当我们强行使用OLS对可观测数据进行回归时：

$$y^1 = \alpha + \beta + \varepsilon, y^0 = \alpha + \varepsilon \quad (9)$$

则由 $\delta = y^1 - y^0 = \beta$ 及 $y = D\delta + y^0$ 可知：

$$y = \alpha + \beta * D + \varepsilon \quad (10)$$

$$\hat{\beta} = ([1, D]'[1, D])^{-1}(1, D)'(y^1, y^0) \quad (11)$$

$$\hat{\alpha} = \bar{Y} - \bar{X}\hat{\beta} \quad (12)$$

由定义可知（假设有 N 个个体）：

$$\begin{aligned} \hat{\alpha} &= \frac{\sum_{D=0} y_i^0}{N_{D=0}} & \hat{\beta} &= \frac{N \left(\frac{\sum_{D=1} y_i^1}{N_{D=1}} \right) - \sum y_i}{N_{D=0}} \\ \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} &= \frac{1}{\mathbf{N}_{D=0}} = \begin{bmatrix} \mathbf{1} & -\mathbf{1} \\ -\mathbf{1} & \frac{\mathbf{N}}{N_{D=1}} \end{bmatrix} \begin{bmatrix} \sum_{i=1, i} y_i^1 \end{bmatrix} \\ \hat{\alpha} &= E[y | D = 0] = \alpha + E[\varepsilon | D = 0] \end{aligned} \quad (13)$$

$$\begin{aligned}
\hat{\beta} &= \frac{N}{N_{D=1}N_{D=0}} \sum_{D=1} y_i^1 - \frac{1}{N_{D=0}} \sum y_i \\
&= \frac{N}{N_{D=1}N_{D=0}} \sum_{D=1} y_i^1 - \frac{N_{D=1}}{N_{D=0}N_{D=1}} \sum y_i^1 - \frac{N_{D=1}}{N_{D=0}N_{D=1}} \sum y_i^0 \\
&= \frac{N_{D=0}}{N_{D=1}N_{D=0}} \sum_{D=1} y_i^1 - \frac{N_{D=1}}{N_{D=0}N_{D=1}} \sum y_i^0 \\
&= \frac{1}{N_{D=1}} \sum_{D=1} y_i^1 - \frac{1}{N_{D=0}} \sum y_i^0
\end{aligned} \tag{14}$$

$$\hat{\beta} = E[y | D = 1] - E[y | D = 0] \tag{15}$$

$$E[y | D = 1] = \alpha + \beta + E[\varepsilon | D = 1] \tag{16}$$

$$E[y | D = 0] = \alpha + E[\varepsilon | D = 0] \tag{17}$$

$$\hat{\beta} = \beta + E[\varepsilon | D = 1] - E[\varepsilon | D = 0] \tag{18}$$

$$E[\hat{\beta}] = E[(D'D)^{-1}(D'y)] = \beta + E[(D'D)^{-1}(D'\varepsilon)] \tag{19}$$

除非D与ε不相关否则为有偏估计. 这里D与ε不相关等价于D与y⁰不相关。

如果使用 Conditional With-Without Comparison (未给定协变量条件) 来估计ATE:

$$\begin{aligned}
E[y_i^1 - y_i^0] &= E_D[E[y_i^1 - y_i^0 | D_i]] \\
E[y_i^1 - y_i^0] &= \pi_i E[y_i^1 - y_i^0 | D_i = 1] + (1 - \pi_i) E[y_i^1 - y_i^0 | D_i = 0] \\
&= E[y_i^1 | D_i = 1] - E[y_i^0 | D_i = 0] + E[y_i^0 | D_i = 0] - E[y_i^0 | D_i = 1] \\
&+ (1 - \pi_i) \{E[y_i^1 | D_i = 0] - E[y_i^1 | D_i = 1] - E[y_i^0 | D_i = 0] + E[y_i^0 | D_i = 1]\} \\
E[y_i^1 | D_i = 1] - E[y_i^0 | D_i = 0] & \\
&= E[y_i^1 - y_i^0] + E[y_i^0 | D_i = 1] \\
&- E[y_i^0 | D_i = 0] + (1 - \pi_i) \{E[y_i^1 | D_i = 1] - E[y_i^0 | D_i = 1] - E[y_i^1 | D_i = 0] + E[y_i^0 | D_i = 0]\}
\end{aligned} \tag{20}$$

上式即可分解为:

$$ATE + selectionbias + (1 - \pi_i)(ATT - ATU) \tag{21}$$

其中最后一项定义为使用平均处理效应的异质性偏差。若各个体同质，则异质性偏差为0，此时仅存在选择偏差。

倾向得分匹配的假设条件

可忽略性假定：给定x_i，则(y_{0i}, y_{1i})独立于D_i，记为：

$$(y_{0i}, y_{1i}) \perp D_i | \mathbf{x}_i \tag{22}$$

倾向得分定理：如果可忽略性假定成立，则在给定p(x)的情况下，(y₀, y₁)独立于D，即：

$$(y_0, y_1) \perp D | \mathbf{x} \Rightarrow (y_0, y_1) \perp D | p(\mathbf{x}) \tag{23}$$

证明：由于D为虚拟变量，故只需证明P[D = 1 | y₀, y₁, p(x)]与y₀, y₁无关即可。

$$\begin{aligned}
& P[D = 1 \mid y_0, y_1, p(\mathbf{x})] \\
&= E[D \mid y_0, y_1, p(\mathbf{x})] \\
&= E_{y_0, y_1, \mathbf{x}} [E(D \mid y_0, y_1, \mathbf{x}) \mid y_0, y_1, p(\mathbf{x})] \quad (\text{迭代期望定律}) \\
&= E_{y_0, y_1, \mathbf{x}} [E(D \mid \mathbf{x}) \mid y_0, y_1, p(\mathbf{x})] \\
&= E_{y_0, y_1, \mathbf{x}} [p(\mathbf{x}) \mid y_0, y_1, p(\mathbf{x})] \\
&= p(\mathbf{x})
\end{aligned} \tag{24}$$

上述证明倒数第二行使用了以下关系式：

$$E(D \mid \mathbf{x}) = 1 \cdot P(D = 1 \mid \mathbf{x}) + 0 \cdot P(D = 0 \mid \mathbf{x}) = p(\mathbf{x}) \tag{25}$$

重叠假定：为了能够进行匹配，需要在 x 的每个可能取值上都同时存在处理组和控制组的个体，即对于 x 的任何可能取值，都有 $0 < p(\mathbf{x}) < 1$ 。

此假定意味着处理组与控制组这两个子样本存在重叠，故名重叠假定，另外，它又是进行匹配的前提，故也称匹配假定，它保证了处理组与控制组的倾向得分取值范围有相同的部分。如果重叠假定不成立，则意味着可能存在某些 x 使得 $p(\mathbf{x}) = 1$ ，即这些个体都属于处理组，无法找到与之匹配的控制组个体；另一方面，也可能存在某些 x 使得 $p(\mathbf{x}) = 0$ ，即这些个体都属于控制组，无法找到与其匹配的处理组个体。因此，在进行匹配时，为了提高匹配的质量，通常仅保留倾向得分重叠部分的个体。

倾向得分匹配的一般步骤

- 1.选择协变量 x_i ：尽量将可能影响 (y_{0i}, y_{1i}) 与 D_i 的相关变量包括进来，以保证可忽略性假设得到满足，如果协变量 x_i 选择不当或太少，导致可忽略性假设不满足，将引起偏差。
- 2.估计倾向得分：一般使用 logit 回归模型估计，Rosenbaum and Rubin(1985) 建议使用形式灵活的logit 回归模型，比如包括 x_i 的高次项与互动项。
- 3.检验平行假定是否满足：如果倾向得分估计得较准确，则应使得 x_i 在匹配后的处理组与控制组之间分布较均匀，比如，匹配后的处理组均值 \bar{x}_{treat} 与控制组均值 $\bar{x}_{control}$ 较接近，这个过程在统计上称为“数据平衡”，但 \bar{x}_{treat} 与 $\bar{x}_{control}$ 的差距显然与计量单位有关，故一般针对 x 的每个分量 x_i 考察如下“标准化差距”：

$$\frac{|\bar{x}_{treat} - \bar{x}_{control}|}{\sqrt{(s_{x, treat}^2 + s_{x, control}^2)/2}} \tag{26}$$

其中， $s_{x, treat}^2$ 与 $s_{x, control}^2$ 分别为处理组与控制组变量 x 的样本方差，一般要求此标准化差距不超过10%，如果超过，则应回到第2步、甚至第1步，重新估计倾向得分。

- 4.进行倾向得分匹配：在进行倾向得分匹配时，常见的匹配方法有以下几种：

近邻匹配法：该方法根据最近的个体进行匹配，然后进行简单算术平均

方法一： k 近邻匹配，即寻找倾向得分最近的 k 个不同组个体，如果 $k = 0$ ，则为一对一匹配，然而，即使“最近邻居”也可能相去甚远，从而失去可比性。

方法二：卡尺匹配或半径匹配，限制倾向得分的绝对距离 $|p_i - p_j| \leq \varepsilon$ ，一般建议 $\varepsilon \leq 0.25\hat{\sigma}_{pscore}$ ，其中 $\hat{\sigma}_{pscore}$ 为倾向得分的样本标准差。

方法三：卡尺内最近邻匹配，即在给定的卡尺范围内寻找最近匹配，此法较为流行。

整体匹配法：每位个体的匹配结果为不同组的全部个体，只是根据个体距离不同给予不同的权重（近者权重大，远者权重小，超出一定范围权重可为0），比如，此时的 \hat{y}_{0i} 估计量为： $\hat{y}_{0i} = \sum_{j:D_j=0} w(i, j)y_j$ ，其中 $w(i, j)$ 为适用于配对 (i, j) 的权重，权重的计算方式不同，又可以分为以下几种方式：

方法一：核匹配，其权重表达式为：

$$w(i, j) = \frac{K[(\mathbf{x}_j - \mathbf{x}_i)/h]}{\sum_{k:D_k=0} K[(\mathbf{x}_k - \mathbf{x}_i)/h]} \quad (27)$$

其中， h 为指定带宽， $K(\cdot)$ 为核函数。

方法二：局部线性回归匹配，即使用局部线性回归来估计权重 $w(i, j)$ 。

方法三：样条匹配，使用更为光滑的“三次样条”来估计权重 $w(i, j)$ 。

在实际进行匹配时，究竟应使用以上哪种具体方法或参数，目前文献中尚无明确指南。一般认为，不存在适用于一切情形的绝对好方法，只能根据具体数据来选择匹配方法。比如，如果控制组个体并不多，则应进行有放回匹配；又比如，如果存在较多具有可比性的控制组个体，则可考虑一对多匹配或核匹配以提高匹配效率。在实践中，一般建议尝试不同的匹配方法，然后比较其结果（类似于敏感度分析），如果不同方法的结果相似，则说明结果是稳健的，不依赖于具体方法，反之，如果存在较大差异，则应考察造成此差异的原因。

5.根据匹配后样本计算平均处理效应：参加者平均处理效应（ATT）估计量的一般表达式为：

$$\widehat{ATT} = \frac{1}{N_1} \sum_{i:D_i=1} (y_i - \hat{y}_{0i}) \quad (28)$$

其中， $N_1 = \sum_i D_i$ 为处理组个数， $\sum_{i:D_i=1}$ 表示仅对处理组个体进行加总。

类似地，也可以为控制组的每位个体 j 寻找处理组的相应匹配，未参加者平均处理效应（ATU）估计量的一般表达式为：

$$\widehat{ATU} = \frac{1}{N_0} \sum_{j:D_j=0} (\hat{y}_{1j} - y_j) \quad (29)$$

其中， $N_0 = \sum_j (1 - D_j)$ 为控制组个体数， $\sum_{j:D_j=0}$ 表示仅对控制组个体进行加总。

整个样本的平均处理效应（ATE）估计量的一般表达式为：

$$\widehat{ATE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_{1i} - \hat{y}_{0i}) \quad (30)$$

其中， $N = N_0 + N_1$ 为样本容量，如果 $D_i = 1$ ，则 $\hat{y}_{1i} = y_i$ ；如果 $D_i = 0$ ，则 $\hat{y}_{0i} = y_i$ 。

倾向得分匹配法的局限性

在一定意义上，匹配估计量可视为一种再抽样方法，因此，在方法论上，倾向得分匹配法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，其思想可以追溯到费舍尔提出的随机实验设计，然而，尽管倾向得分匹配法可能在很大程度上减少观测数据的偏差，但它本身也有如下局限性：

1.倾向得分匹配通常要求比较大的样本容量以得到高质量的匹配；

2.倾向得分匹配要求处理组与控制组的倾向得分有较大的共同取值范围，否则，将丢失较多观测值，导致剩下的样本不具有代表性；

3.倾向得分匹配只控制了可测变量的影响，如果存在依不可测变量选择，仍会带来隐性偏差。

倾向得分匹配的实例

以参加职业培训是否能提高实际收入作为例子，来介绍倾向得分匹配方法的具体应用。以数据集 `ldw_exper.dta` 为例进行演示，该数据集由 Dehejia and Wahba(1999) 构建，为 Abadie et al(2004) 所使用，原始数据来自 Lalonde(1986)。该数据集包括以下变量：结果变量 `re78` (1978年实际收入)，处理变量 `t` (是否参加就业培训)，协变量 `age` (年龄)，`educ` (教育年限)，`black` (是否黑人)，`hisp` (是否拉丁裔)，`married` (是否已婚)，`re74` (1974年实际收入)，`re75` (1975年实际收入)，`u74` (1974年是否失业)以及`u75`(1975年是否失业)。

```
*安装命令
ssc install psmatch2,replace
```

```
*导入数据
use ldw_exper.dta,clear //找不到上面数据的话可以通过help ldw_exper.dta命令获取
```

```
*进行倾向得分匹配，先将数据随机排序
set seed 10101
gen ranorder = runiform()
sort ranorder
```

```
*1.一对一匹配，由于样本容量并不大，进行有放回的匹配，且允许并列
psmatch2 t age educ black hisp married re74 re75 u74 u75,outcome(re78) n(1) ate ties
logit common
*考虑使用自助法得到标准误
set seed 10101
bootstrap r(att) r(atu) r(ate),reps(500):psmatch2 t age educ black hisp married re74
re75 u74 u75,outcome(re78) n(1) ate ties logit common
```

一对一匹配结果：

```
Logistic regression                Number of obs    =        445
                                   LR chi2(9)          =        11.70
                                   Prob > chi2         =        0.2308
Log likelihood = -296.25026         Pseudo R2       =        0.0194
```

t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0142619	.0142116	1.00	0.316	-.0135923	.0421162
educ	.0499776	.0564116	0.89	0.376	-.060587	.1605423
black	-.347664	.3606532	-0.96	0.335	-1.054531	.3592032
hisp	-.928485	.50661	-1.83	0.067	-1.921422	.0644523

married		.1760431	.2748817	0.64	0.522	-.3627151	.7148012
re74		-.0339278	.0292559	-1.16	0.246	-.0912683	.0234127
re75		.01221	.0471351	0.26	0.796	-.0801731	.1045932
u74		-.1516037	.3716369	-0.41	0.683	-.8799987	.5767913
u75		-.3719486	.317728	-1.17	0.242	-.9946841	.2507869
_cons		-.4736308	.8244205	-0.57	0.566	-2.089465	1.142204

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
re78	Unmatched	6.34914538	4.55480228	1.79434311	.632853552	2.84
	ATT	6.40495818	4.99436488	1.4105933	.839875971	1.68
	ATU	4.52683013	6.15618973	1.6293596	.	.
	ATE			1.53668776	.	.

Note: S.E. does not take into account that the propensity score is estimated.

```

psmatch2: | psmatch2: Common
Treatment | support
assignment | Off suppo On suppor | Total
-----+-----+-----
Untreated | 11 249 | 260
Treated | 2 183 | 185
-----+-----+-----
Total | 13 432 | 445

```

使用自助法得到标准误:

```

Bootstrap results           Number of obs   =       445
                             Replications       =       500

```

```

command: psmatch2 t age educ black hisp married re74 re75 u74 u75, outcome(re78)
n(1) ate ties logit common
      _bs_1: r(att)
      _bs_2: r(atu)
      _bs_3: r(ate)

```

	Observed	Bootstrap			Normal-based	
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_bs_1	1.410593	.8849862	1.59	0.111	-.3239478 3.145134	
_bs_2	1.62936	.9416177	1.73	0.084	-.2161772 3.474896	
_bs_3	1.536688	.7358839	2.09	0.037	.0943818 2.978994	

*使用pstest来考察此匹配结果是否较好地平衡了数据

```

quietly psmatch2 t age educ black hisp married re74 re75 u74 u75,outcome(re78) n(1) ate
ties logit common
pstest age educ black hisp married re74 re75 u74 u75,both graph

```

Variable	Unmatched Matched	Mean		%reduct %bias	t-test		
		Treated	Control		bias	t	p> t
age	U	25.816	25.054	10.7		1.12	0.265
	M	25.781	25.383	5.6	47.7	0.52	0.604
educ	U	10.346	10.088	14.1		1.50	0.135
	M	10.322	10.415	-5.1	63.9	-0.49	0.627
black	U	.84324	.82692	4.4		0.45	0.649
	M	.85246	.86339	-2.9	33.0	-0.30	0.765
hisp	U	.05946	.10769	-17.5		-1.78	0.076
	M	.06011	.04372	5.9	66.0	0.71	0.481
married	U	.18919	.15385	9.4		0.98	0.327
	M	.18579	.19126	-1.4	84.5	-0.13	0.894
re74	U	2.0956	2.107	-0.2		-0.02	0.982
	M	2.0672	1.9222	2.7	-1166.6	0.27	0.784
re75	U	1.5321	1.2669	8.4		0.87	0.382
	M	1.5299	1.6446	-3.6	56.7	-0.32	0.748
u74	U	.70811	.75	-9.4		-0.98	0.326
	M	.71038	.75956	-11.1	-17.4	-1.06	0.288
u75	U	.6	.68462	-17.7		-1.85	0.065
	M	.60656	.63388	-5.7	67.7	-0.54	0.591

上表显示，匹配后大多数变量的标准化偏差% bias小于10%，只是变量u74 的偏差为11.1%，似乎可以接受，而且大多数t检验的结果不拒绝处理组与控制组无系统差异的原假设，对比匹配前的结果，大多数变量的标准化偏差均大幅缩小。各变量标准化偏差的匹配前后变化参见图1

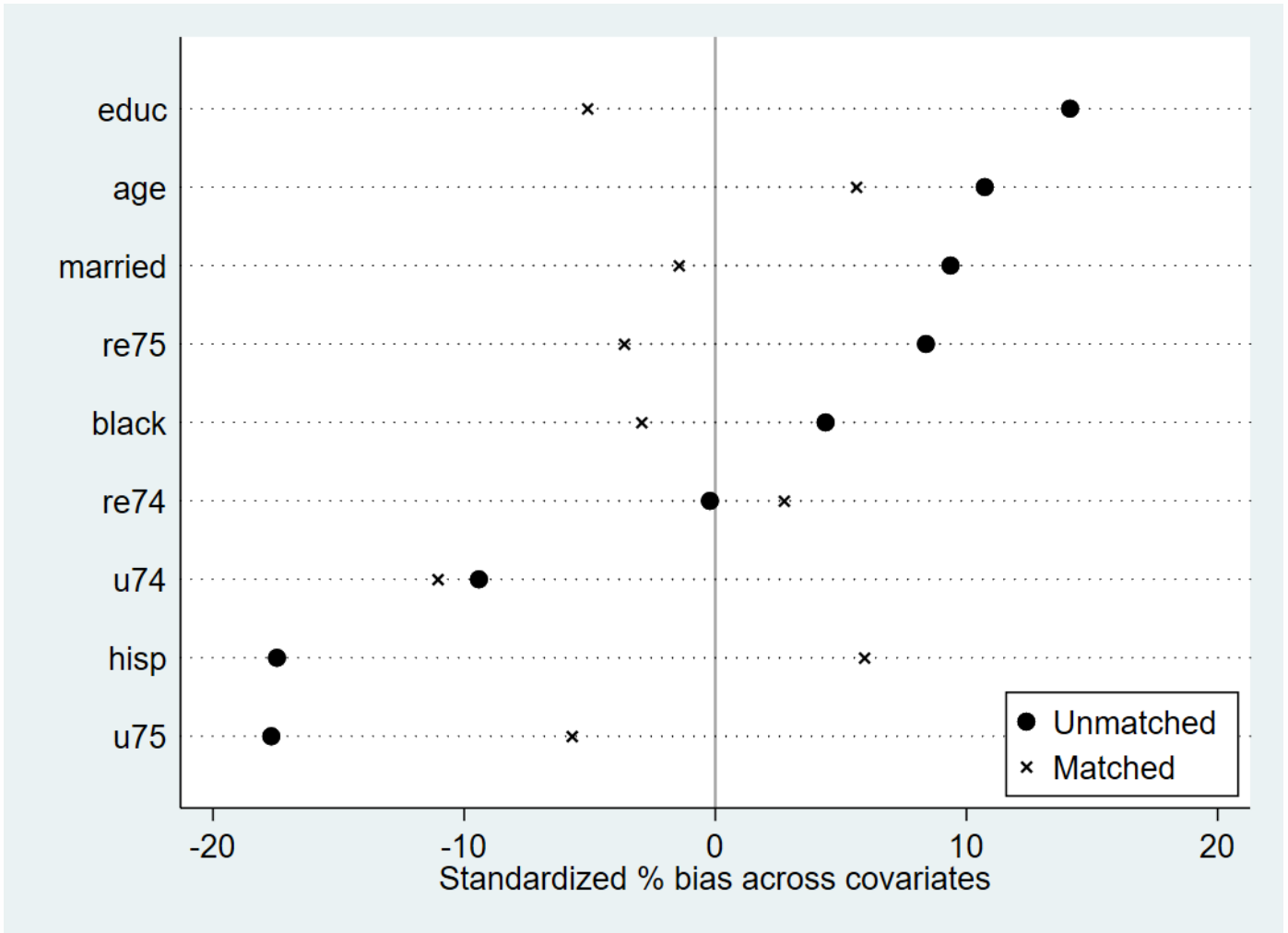


图1 各变量标准化偏差图示

*画条形图来显示倾向得分的共同取值范围
psgraph

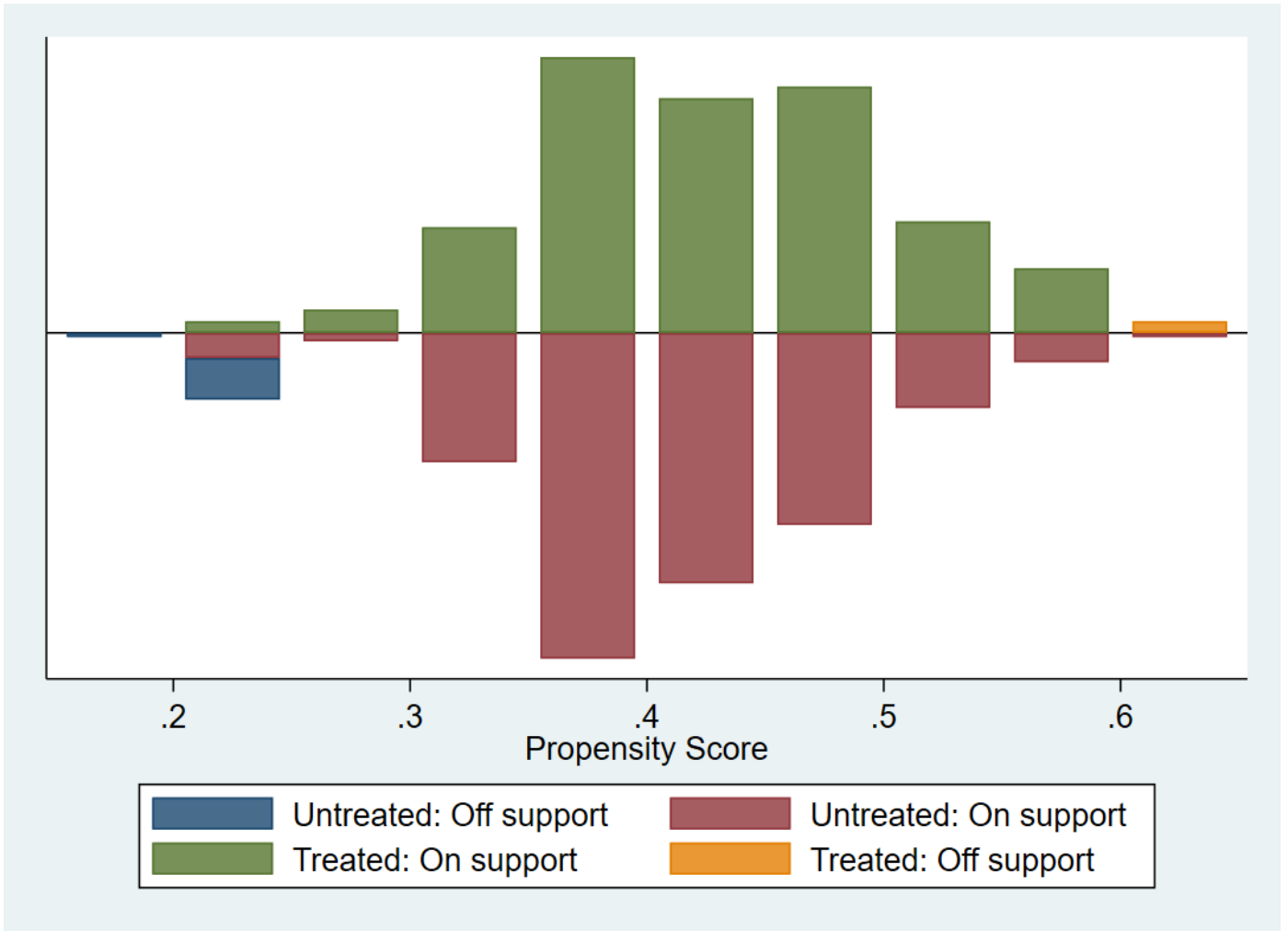


图2 倾向得分的共同取值范围

从图2可以直观地看出，大多数观测值在共同取值范围内(*Onsupport*)，故在进行倾向得分匹配时仅会损失少了样本。

*2. 近邻匹配，令k=4，为节省空间，使用选项quietly略去对倾向得分估计结果的汇报

```
psmatch2 t age educ black hisp married re74 re75 u74 u75,outcome(re78) n(4) ate ties
logit common quietly
```

近邻匹配结果:

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
re78	Unmatched	6.34914538	4.55480228	1.79434311	.632853552	2.84
	ATT	6.40495818	4.3883555	2.01660267	.727464025	2.77
	ATU	4.52683013	5.74351061	1.21668048	.	.
	ATE			1.55553641	.	.

Note: S.E. does not take into account that the propensity score is estimated.

```
psmatch2: | psmatch2: Common
Treatment | support
assignment | Off suppo On suppor | Total
-----+-----+-----
Untreated | 11 249 | 260
```

Treated		2	183		185
-----+-----+-----					
Total		13	432		445

从上述结果可知，一对四匹配的结果与一对一匹配类似，只是(ATT)的估计值有较大差异，且在5%水平上显著。

```
*3.卡尺内一对四匹配，首先计算倾向得分（由Stata自动生成并记录于变量_pscore）的标准差，然后乘0.25
sum _pscore
dis 0.25 * r(sd) //计算结果约为0.02，为保守起见，将卡尺范围定为0.01，意味着对倾向得分相差1%的观测值
进行一对四匹配
psmatch2 t age educ black hisp married re74 re75 u74 u75,outcome(re78) n(4) cal(0.01)
ate ties logit common quietly
```

卡尺内一对四匹配结果：

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
re78	Unmatched	6.34914538	4.55480228	1.79434311	.632853552	2.84
	ATT	6.40495818	4.35102383	2.05393435	.729511502	2.82
	ATU	4.51395698	5.63458891	1.12063193	.	.
	ATE			1.52061868	.	.

Note: S.E. does not take into account that the propensity score is estimated.

```
psmatch2: | psmatch2: Common
Treatment | support
assignment | Off suppo On suppor | Total
-----+-----+-----
Untreated | 16 244 | 260
Treated | 2 183 | 185
-----+-----+-----
Total | 18 427 | 445
```

上表显示，卡尺内一对四匹配结果与简单的一对四匹配比较接近，这说明大多数一对四匹配均发生在卡尺0.01的范围内，不存在太远的近邻。

```
*4.半径（卡尺）匹配
psmatch2 t age educ black hisp married re74 re75 u74 u75,outcome(re78) radius cal(0.01)
ate ties logit common quietly
```

半径（卡尺）匹配结果：

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
re78	Unmatched	6.34914538	4.55480228	1.79434311	.632853552	2.84
	ATT	6.40495818	4.54598746	1.85897071	.704721737	2.64
	ATU	4.51395698	6.14730708	1.63335009	.	.
	ATE			1.73004464	.	.

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2:		psmatch2: Common		
Treatment	support			
assignment	Off suppo	On suppor		Total
Untreated	16	244		260
Treated	2	183		185
Total	18	427		445

此匹配结果依然类似。

*5.核匹配 (使用默认的核函数与带宽)

```
psmatch2 t age educ black hisp married re74 re75 u74 u75,outcome(re78) kernel ate ties
logit common quietly
```

核匹配结果:

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
re78	Unmatched	6.34914538	4.55480228	1.79434311	.632853552	2.84
	ATT	6.40495818	4.58467178	1.82028639	.685942087	2.65
	ATU	4.52683013	6.13018214	1.603352	.	.
	ATE			1.69524782	.	.

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2:		psmatch2: Common		
Treatment	support			
assignment	Off suppo	On suppor		Total
Untreated	11	249		260
Treated	2	183		185
Total	13	432		445

结果依然类似。

*6.局部线性回归匹配 (使用默认的核函数和带宽)

```
psmatch2 t age educ black hisp married re74 re75 u74 u75,outcome(re78) llr ate ties
logit common quietly
```

*上面未汇报ATT的标准误, 为此使用自助法得到自助标准误

```
set seed 10101
```

```
bootstrap r(att) r(atu) r(ate),reps(500):psmatch2 t age educ black hisp married re74
re75 u74 u75,outcome(re78) llr ate ties logit common quietly
```

局部线性回归匹配结果:

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
----------	--------	---------	----------	------------	------	--------

```

re78 Unmatched | 6.34914538 4.55480228 1.79434311 .632853552 2.84
      ATT | 6.40495818 4.71399307 1.69096511 .839875971 2.01
      ATU | 4.52683013 6.20121299 1.67438286 . .
      ATE | . 1.68140728 . .

```

Note: S.E. does not take into account that the propensity score is estimated.

```

psmatch2: | psmatch2: Common
Treatment | support
assignment | Off suppo On suppor | Total
-----+-----+-----+-----
Untreated | 11 249 | 260
Treated | 2 183 | 185
-----+-----+-----+-----
Total | 13 432 | 445

```

*使用自助法得到自助标准误:

```

Bootstrap results          Number of obs   =      445
                          Replications       =      500

```

```

command: psmatch2 t age educ black hisp married re74 re75 u74 u75, outcome(re78) llr
ate ties logit common quietly
      _bs_1: r(att)
      _bs_2: r(atu)
      _bs_3: r(ate)

```

	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]
_bs_1	1.690965	.7152386	2.36	0.018	.2891232 3.092807
_bs_2	1.674383	.6928838	2.42	0.016	.3163555 3.03241
_bs_3	1.681407	.666897	2.52	0.012	.3743132 2.988501

根据上表的自助标准误，对平均处理效应的三助度量均至少在5%水平上显著。

```

*7.样条匹配（同样使用自助法计算标准误），先安装一个非官方命令spline
findit snp7_1
set seed 10101
bootstrap r(att) r(atu) r(ate),reps(500):psmatch2 t age educ black hisp married re74
re75 u74 u75,outcome(re78) spline ate ties logit common quietly

```

样条匹配结果:

```

Bootstrap results          Number of obs   =      445
                          Replications       =      500

```

```

command: psmatch2 t age educ black hisp married re74 re75 u74 u75, outcome(re78)
spline ate ties logit common quietly
      _bs_1: r(att)

```

```
_bs_2: r(atu)
_bs_3: r(ate)
```

	Observed	Bootstrap			Normal-based	
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_bs_1	1.74968	.7024615	2.49	0.013	.3728811	3.126479
_bs_2	1.576453	.6871374	2.29	0.022	.2296887	2.923218
_bs_3	1.649834	.6559297	2.52	0.012	.3642357	2.935433

估计结果仍然类似，总之，以上各种倾向得分匹配的结果表明，参加就业培训的平均处理效应为正，不仅在经济上显著，而且在统计上显著。

*8. 马氏匹配，并使用Adadie and Imbens(2006)提供的异方差稳健标准误

```
psmatch2 t,outcome(re78) mahal(age educ black hisp married re74 re75 u74 u75) n(4)
ai(4) ate
```

马氏匹配结果：

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
re78	Unmatched	6.34914538	4.55480228	1.79434311	.632853552	2.84
	ATT	6.34914538	4.45101599	1.89812939	.707602907	2.68
	ATU	4.55480228	6.40808613	1.85328385	.902132479	2.05
	ATE			1.8719275	.784605816	2.39

Note: Sample S.E.

psmatch2:		
psmatch2:	Common	
Treatment	support	
assignment	On suppor	Total
Untreated	260	260
Treated	185	185
Total	445	445

上表显示，无论是平均处理效应的估计值还是显著性，马氏匹配结果与倾向得分匹配类似。

大样本下的倾向得分匹配与修正

设反事实模型：

$$\begin{aligned} y_i &= D_i f^1(x_i) + (1 - D_i) f^0(x_i) + \varepsilon_i \\ \hat{\tau}_{ATT} &= \frac{1}{N_{D=1}} \sum_{D=1} [f^1(x_i) + \varepsilon_i - (f^0(x_{j(i)}) + \varepsilon_{j(i)})] \\ &= \frac{1}{N_{D=1}} \sum_{D=1} [f^1(x_i) - f^0(x_{j(i)}) + \varepsilon_{j(i)} - \varepsilon_i] \\ &= \frac{1}{N_{D=1}} \sum_{D=1} [f^1(x_i) - f^0(x_i) + f^0(x_i) - f^0(x_{j(i)}) + \varepsilon_{j(i)} - \varepsilon_i] \end{aligned} \quad (31)$$

因此，

$$\hat{\delta}_{ATT} - \delta_{ATT} = \frac{1}{N_{D=1}} \sum_{D=1} [f^1(x_i) - f^0(x_i) - \delta_{ATT} + f^0(x_i) - f^0(x_{j(i)}) + \varepsilon_{j(i)} - \varepsilon_i] \quad (32)$$

由大数定律, 其中前三项和最后两项在 $N_{D=1} \rightarrow \infty$ 时, 乘以 $\sqrt{N_{D=1}}$ 后, 各自满足均值为零的正态分布, 但剩余的中间两项为:

$$\frac{1}{\sqrt{N_{D=1}}} \sum_{D=1} [f^0(x_i) - f^0(x_{j(i)})] \xrightarrow{P} E[f^0(x_i) - f^0(x_{j(i)}) | D_i = 1] \quad (33)$$

因此, 定义偏差校正估计量 (Bias Correction Estimator) 为:

$$\hat{\delta}_{ATT}^{BC} := \hat{\delta}_{ATT} - \frac{1}{N_{D=1}} \sum_{D=1} (\hat{f}^0(x_i) - \hat{f}^0(x_{j(i)})) \quad (34)$$

其中, \hat{f}^0 可由对线性回归得到。

概率倒数权重法

使用倾向得分方法虽然消弭了个体做出选择的特征性偏差, 但它也可能会损失效率。Hahn(1998)和 Heckman, Ichimura和Todd(1998)表明, 相比于对预处理变量一一匹配, 通过倾向得分匹配会导致效率损失。然而, Rosenbaum(1987)和Rubin, Thomas(1997)证明, 使用倾向得分的参数估计, 而不是实际的倾向得分 (true selection probability), 可以避免部分效率损失。Rotnitzky和Robins(1995)在存在缺失数据的回归模型中提出了类似的观点, 其中缺失数据是随机缺失的 (之后推导部分将选择概率设定为0.5亦由此参考)。这些文章表明, 使用个体被选择概率的参数估计的逆作为加权比用真实的选择概率更有效。

通过逆概率加权法得出平均处理效应的估计往往比直接匹配得到的估计要更为有效, 原因在于逆概率加权解释了选择的非随机性——我们所得到的观察个体非随机选择组别, 其匹配的对照组个体便会非随机缺失, 因而造成匹配缺失和选择偏差, 这种偏差可以通过逆概率加权法纠正。注意到:

$$ATE = | (E[y | x, D = 1] - E[y | x, D = 0]) dP_x \quad (35)$$

被积差项以confounder x 为条件分层. 虽然在各层内实现了对 x 的控制, 结合CIA实现了随机试验, 但 x 在 treated 和 untreated group 中的分布情况很可能不同(imbalance), 进而需要用统一的分布来计算均值意义上的处理效果. 即通过特相应re-weighting纠正试验控制两组特征分布差异带来的影响。

理论模型

首先，考虑一个简单的模型，我们从大量人口中随机得到样本容量为N的样本。对于样本中的每个个体i，令T表示是否接受了感兴趣的处理，如果单元i选择了处理组， $T_i = 1$ ，如果个体i选择了对照处理， $T_i = 0$ 。

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0) \quad (36)$$

在此条件下，当给定协变量x时平均处理效应即为 $\tau(x)$

$$\tau(x) \equiv E[Y(1) - Y(0) | X = x] \quad (37)$$

$$\begin{aligned} & E[Y(1) - Y(0) | X = x] \\ &= E[Y(1) | T = 1, X = x] - E[Y(0) | T = 0, X = x] \\ &= E[Y | T = 1, X = x] - E[Y | T = 0, X = x] \end{aligned} \quad (38)$$

而在实际情况中，如果x包含的元素过多，并且在这种情况下，我们对于处理组和控制组的匹配要求完全相同，那么我们可能并不能得到想要的结果。为了避免不得不根据所有预处理变量的值来在个体之间进行匹配，Rosenbaum和Rubin(1983, 1984)开发了一种基于倾向评分的方法，即选择个体进入处理组的概率（后也表述为 $p(x)$ ）：

$$e(x) = \Pr(T = 1 | X = x) \quad (39)$$

我们假设这个概率是有界的(0,1)。倾向匹配的核心思想在于，如果进行政策冲击和潜在结果独立于所有预处理变量，那么两者也将独立于给定预处理变量的接受治疗的条件概率，即独立于倾向评分的条件概率。

$$T \perp (Y(0), Y(1)) | e(X) \quad (40)$$

之后，由Horvitz Thompson(1952)提出对于PSM的一种有效替代方法是，通过个体选择概率的逆来重新加权观察值，并将其加权平均作为治疗效果的估计。这个加权估计量可以写成：

$$\hat{\tau}_{ht} = \frac{1}{N} \sum_{i=1}^N \left[\frac{t_i \cdot y_i}{e(x_i)} - \frac{(1 - t_i) \cdot y_i}{1 - e(x_i)} \right] \quad (41)$$

关于上式，我们可以看到，各项分母为其选择概率，当 $t=1$ 时，仅保留第一项；当 $t=0$ 时，仅保留第2项，进一步，我们可以将上式简化为：

根据非混淆性假设，我们可以得到：

$$\begin{aligned} & E[y^1] \\ &= E_x [E[y^1 | x]] \\ &= E_x \left[\frac{E[y^1 | x] E[D | x]}{E[D | x]} \right] \\ &= CIA E_x \left[\frac{E[D \cdot y^1 | x]}{p(x)} \right] \\ &= E_x \left[\frac{E[D \cdot y | x]}{p(x)} \right] \\ &= E_x \left[E \left[\frac{D \cdot y}{p(x)} | x \right] \right] \\ &= E \left[\frac{D \cdot y}{p(x)} \right] \end{aligned} \quad (42)$$

$$\begin{aligned}
E[y^0] &= E_x[E[y^0 | x]] \\
E_x \left[E \left[\frac{E[y^0 | x]}{1 - E[D | -D | x | x]} \right] \right] &= E \left[\frac{(1 - D) \cdot y}{1 - p(x)} \right]
\end{aligned} \tag{43}$$

或者，当我们考虑x为连续变量时：

$$\begin{aligned}
&E[y^1] \\
&= E_x[E[y^1 | x]] \\
&=^{CIA} E_x[E[y^1 | D = 1, x]] \\
&= \int_x \int_{y^1} y^1 dF_{y^1|D,x}(D = 1) dF_x \\
&= \int_x \int_{y^1} y^1 \frac{dF_{D|x}(D = 1)}{dF_{D|x}(D = 1)} dF_{y^1|D,x}(D = 1) dF_x \\
&= \int_x \int_{y^1} \frac{y^1}{dF_{D|x}(D = 1)} dF_{y^1,D,x}(D = 1) \\
&= \int_D \int_x \int_{y^1} \frac{\mathbf{1}_{\{D=1\}} y^1}{dF_{D|x}} dF_{y^1,D,x} \\
&= E \left[\frac{\mathbf{1}_{\{D=1\}} y^1}{dF_{D|x}} \right] \\
&= E \left[\frac{D \cdot y}{p(x)} \right]
\end{aligned} \tag{44}$$

因此，

$$\tau_{ATE} = E \left[\frac{y \cdot (D - p(x))}{p(x)(1 - p(x))} \right] \tag{45}$$

$$\hat{\tau}_{ATE} = \frac{1}{N} \sum_{i=1}^N \left[\frac{Y_i \cdot (D - \hat{p}(x_i))}{\hat{p}(x_i)(1 - \hat{p}(x_i))} \right] \tag{46}$$

上式即为使用逆概率加权的ATE估计。

对于ATT，考察：

$$E[y^0] = E[y^0 | D = 1] \Pr(D = 1) + E[y^0 | D = 0] \Pr(D = 0) \tag{47}$$

其中：

$$\begin{aligned}
&E[y^0 | D = 0] \Pr(D = 0) \\
&= \int_{y^0} y^0 dF_{y^0|D}(D = 0) \int_D \mathbf{1}_{\{D=0\}} dF_D \\
&= \int_{y^0} \mathbf{1}_{\{D=0\}} y^0 dF_{y^0|D} \int_D \mathbf{1}_{\{D=0\}} dF_D \\
&= \int_D \int_{y^0} \mathbf{1}_{\{D=0\}} y^0 dF_{y^0,D} \\
&= \int_D \int_{y^0} (1 - D) y^0 dF_{y^0,D} \\
&= E[(1 - D) y^0] \\
&= E[(1 - D) y]
\end{aligned} \tag{48}$$

同理，

$$\begin{aligned}
& E[y^1 | D = 1] \Pr(D = 1) \\
&= \int_{y^1} y^1 dF_{y^1|D}(D = 1) \int_D \mathbf{1}_{\{D=1\}} dF_D \\
&= \int_{y^1} \mathbf{1}_{\{D=1\}} y^1 dF_{y^1|D} \int_D \mathbf{1}_{\{D=1\}} dF_D \\
& \int_D \int_{y^1} \mathbf{1}_{\{D=1\}} y^1 dF_{y^1,D} = \int_D \int_{y^1} Dy^1 dF_{y^1,D} = E[Dy^1] = E[Dy]
\end{aligned} \tag{49}$$

因此，

$$E[y^0 | D = 1] = \frac{E(y^0) - E[(1 - D)y]}{\Pr(D = 1)} \tag{50}$$

注：此处D=1的概率非实际概率，实际概率的效率较低，更有效的方法是使用非参估计值。

因此，

$$\begin{aligned}
\delta_{ATT} &= E[y^1 - y^0 | D = 1] = \frac{E[Dy] - E[y^0] + E[(1 - D)y]}{\Pr(D = 1)} \\
&= \frac{E[y] - E\left[\frac{(1-D) \cdot y}{1-p(x)}\right]}{\Pr(D = 1)} = \frac{1}{\Pr(D = 1)} \left(E[y] - E\left[\frac{(1 - D) \cdot y}{1 - p(x)}\right] \right)
\end{aligned} \tag{51}$$

$$\begin{aligned}
&= \frac{1}{\Pr(D = 1)} E\left[\frac{(D - p(x)) \cdot y}{1 - p(x)}\right] \\
\delta_{ATT} &= \frac{1}{\Pr(D = 1)} E\left[\frac{(D - p(x)) \cdot y}{1 - p(x)}\right]
\end{aligned} \tag{52}$$

$$\hat{\delta}_{ATT} = \frac{1}{N_{D=1}} \sum_{i=1}^N Y_i \cdot \frac{D_i - \hat{p}(x_i)}{1 - \hat{p}(x_i)} \tag{53}$$

IPTW的有效性

接下来，证明逆概率加权法的有效性，为简化起见，考虑x为离散值时的情况，假设只存在一个二元的协变量（x=0或x=1），并将 N_{tx} 设为当t={0,1}、x={0,1}时的观测个体数，因此存在：

$$N_{\cdot x} = N_{0x} + N_{1x} \tag{54}$$

在这个模型中，我们将倾向得分等价于被选择概率（当两者被选择概率近似时，这也意味着此类个体更有可能进入同一组，与之匹配的相对应组别的类似个体会更少）

$$p(x) = E[T | X = x] = \Pr(T = 1 | X) \tag{55}$$

进一步地，假设无论对于x=1还是x=0的个体而言，选择概率为常数值0.5，则样本中的缺失值完全是随机的，设 $\beta_0 = E[Y]$ ， β_0 的标准化方差界为：

$$V_{\text{bound}} = 2 \cdot E[V(Y | X)] + V(E[Y | X]) \tag{56}$$

接着，考虑四种情况下的 β_0 的估计值

第一种情况，通过样本观测值的均值计算总体平均效应

$$\hat{\beta}_{\text{comp}} = \frac{\sum_{i=1}^N y_i \cdot t_i}{\sum_{j=1}^N t_j} \tag{57}$$

由于我们所能观测到的存在50%的缺失值，将0.5带入，因此 β_0 的标准差为：

$$V_{\text{comp}} = 2 \cdot E[V(Y | X)] + 2 \cdot V(E[Y | X]) = 2V(Y) \quad (58)$$

第二种情况，通过实际权重计算 β_0 的估计值（将实际ps取逆）：

$$\hat{\beta}_{tw} = \frac{1}{N} \sum_{i=1}^N \frac{y_i \cdot t_i}{p(x_i)} = \frac{1}{N} \sum_{i=1}^N \frac{y_i \cdot t_i}{1/2} \quad (59)$$

其大样本情况下的标准方差为（大于情况1的方差）：

$$V_{tw} = 2 \cdot E[V(Y | X)] + V(E[Y | X]) + E[E[Y | X]^2] = 2V(Y) + E(Y^2) \quad (60)$$

接下来，第三种估计量通过ps的非参数估计的逆对可观测值进行加权平均。这种估计值是IPTW的核心思想，在当前的背景下，我们所估计的选择概率只是一个协变量的给定值(0,1)的可观察数量的比例。因此，估计的个体被选择概率为：

$$\hat{p}(x) = \begin{cases} N_{10}/N_{\cdot 0} & \text{if } x = 0 \\ N_{11}/N_{\cdot 1} & \text{if } x = 1 \end{cases} \quad (61)$$

因此，在这种情况下的“估计权重”的 β_0 的估计值为：

$$\hat{\beta}_{ew} = \frac{1}{N} \sum_{i=1}^N \frac{y_i \cdot t_i}{\hat{p}(x_i)} \quad (62)$$

其标准化方差与方差界相等，估计量的效率提高。

$$V_{ew} = 2 \cdot E[V(Y | X)] + V(E[Y | X]) = V_{\text{bound}} \quad (63)$$

因此，在这种简单情况下，具有非参匹配权重的加权估计值不仅比使用“真”权重的估计值具有更低的方差，而且在达到方差界的意义上更有效。接下来我们将说明为什么这种估计值在协变量为连续时依然有效。

第四种情况是通过对x进行非参估计由此计算 β_0 的估计值。

$$\hat{\mu}(x) = \sum_{i|x_i=x} y_i \cdot t_i / \sum_{i|x_i=x} t_i \quad (64)$$

上式为x=0或1下的条件回归下的非参估计量，即：

$$\mu(x) = E[Y | T = 1, X = x] \quad (65)$$

接着考虑在这种情况下的 β_0 的条件估计值：

$$\hat{\beta}_{rc} = \frac{1}{N} \sum_{i=1}^N \hat{\mu}(x_i) \quad (66)$$

考虑 $\hat{\beta}_{rc}$ 的构建思路，当t=1时，使用可得的观测数据；当t=0时，使用估计值

$$\hat{\beta}_{rc} = \frac{1}{N} \sum_{i=1}^N t_i \cdot y_i + (1 - t_i) \cdot \hat{\mu}(x_i) \quad (67)$$

Hahn(1998)在其论文中论证了第三种情况与第二种情况的估计结果是一致且有效的。

在当前条件下，协变量仅为二元变量，因此，我们可以将 $\hat{\beta}_{rc}$ 重写为：

$$\begin{aligned}
\hat{\beta}_{rc} &= \frac{1}{N} [(N_{00} + N_{10}) \cdot \hat{\mu}(0) + (N_{01} + N_{11}) \cdot \hat{\mu}(1)] \\
&= \frac{1}{N} \left[(N_{00} + N_{10}) \cdot \frac{\sum_{i: x_i=0} t_i \cdot y_i}{N_{10}} + (N_{01} + N_{11}) \cdot \frac{\sum_{i: x_i=1} t_i \cdot y_i}{N_{11}} \right] \\
&= \frac{1}{N} \sum_i \frac{t_i \cdot y_i}{w_i}
\end{aligned} \tag{68}$$

其中,

$$w_i = \begin{cases} N_{10}/N_{.0} & \text{if } x_i = 0 \\ N_{11}/N_{.1} & \text{if } x_i = 1, \end{cases} \tag{69}$$

在第四种情况下, w_i 完全等同于 $\hat{p}(x_i)$, 因此, 两种情况下的 β_0 的估计值完全相同——在离散预处理变量条件下, 非参选择概率的估计量与非参权重的估计量在数学含义上完全等同, 这说明非参权重估计量在离散预处理协变量条件下也是有效估计值。

根据我们在第一部分给出的倾向得分计算公式, 上述条件也可被改写为:

$$\Pr(T = 1 | X = x) = \frac{\exp(\delta \cdot (\hat{\mu}(x) - \hat{\beta}))}{1 + \exp(\delta \cdot (\hat{\mu}(x) - \hat{\beta}))} \tag{70}$$

$$w_i = \frac{\exp(\hat{\delta} \cdot (\hat{\mu}(x_i) - \hat{\beta}))}{1 + \exp(\hat{\delta} \cdot (\hat{\mu}(x_i) - \hat{\beta}))} \tag{71}$$

$$\hat{\beta}_{rr} = \frac{1}{N} \sum_{i=1}^N \frac{y_i \cdot t_i}{w_i} \tag{72}$$

由于在第一种情况中, 我们已经讨论过方差效率界的计算并不依赖于协变量是否离散这一条件, 据此, Chamberlain (1987)使用GMM方法重新证明了第三类、第四类估计量的有效性 (在协变量为连续变量时)。

在此模型中, 我们仍然使用 $p=0.5$ 的假设, 使用一阶矩条件: $E[\psi_1(Y, X, T, \beta_0)] = 0$, 则:

$$\psi_1(y, t, x, \beta) = y \cdot t / p(x) - \beta = \frac{y \cdot t}{1/2} - \beta \tag{73}$$

在上述条件下我们所计算的是“真实权重”取逆下的估计值 $\hat{\beta}_{tw}$. 然而这种情况并非是有效的, 因为它忽略了以选择概率形式透露出的附加信息:

$$\begin{aligned}
E[T | X = x] &= p(x) = 1/2 \\
E[T - 1/2 | X] &= 0
\end{aligned} \tag{74}$$

根据式 (74), 假设 $x=(0,1)$, 我们可以重写一阶矩条件 $E[\psi_2(Y, X, T, \beta_0)] = 0$ 为

$$\psi_2(y, t, x, \beta) = \begin{pmatrix} x \cdot (t - 1/2) \\ (1 - x) \cdot (t - 1/2) \end{pmatrix} \tag{75}$$

在GMM中, β_0 的估计值依据上式即可得到有效估计量 (通过矩条件)。同样的, 通过极大似然函数也可求解:

$$L(\pi) = \sum_{i=1}^N \ln \pi_i \tag{76}$$

这一极大似然函数在下属条件下成立

$$\sum_{i=1}^N \pi_i = 1 \tag{77}$$

满足一阶矩条件 $\psi_1(\cdot)$:

$$\sum_{i=1}^N \pi_i \cdot \left(\frac{y_i \cdot t_i}{1/2} - \beta \right) = 0 \quad (78)$$

满足个体被选择概率中所包含的x的信息 $\psi_2(\cdot)$:

$$\sum_{i=1}^N \pi_i \cdot x_i \cdot (t_i - 1/2) = 0 \quad (79)$$

及

$$\sum_{i=1}^N \pi_i \cdot (1 - x_i) \cdot (t_i - 1/2) = 0 \quad (80)$$

根据上述条件构造拉格朗日函数计算得到 $\hat{\pi}_i$ ，并带入第四种情况下的 β 的估计值可知，

$$\hat{\beta}_{el} = \frac{\sum_{i=1}^N \hat{\pi}_i \cdot y_i \cdot (2t_i)}{\sum_{i=1}^N \hat{\pi}_i} = \sum_{i=1}^N 2\hat{\pi}_i \cdot y_i \cdot t_i \quad (81)$$

其中， $\hat{\pi}_i$ 可求解为：

$$\hat{\pi}_i = \left(1 + \frac{N \cdot 11/N_1 - 1/2}{1/4} \cdot x_i \cdot (t_i - 1/2) + \frac{N_{10}/N_0 - 1/2}{1/4} \cdot (1 - x_i) \cdot (t_i - 1/2) \right)^{-1} \quad (82)$$

因此，带入式 (81) 可知

$$\hat{\beta}_{el} = \sum_{i=1}^N 2\hat{\pi}_i \cdot y_i \cdot t_i = \hat{\beta}_{ew} \quad (83)$$

因此，该估计量是有效的。

综上所述，使用倾向评分的非参数估计的逆对观察值加权平均能得到更为有效的估计。

参考文献

- [1] 陈强.高级计量经济学及Stata应用.2版.北京:高等教育出版社,2014.
- [2] Dehejia ,R.H., and S. Wahba , 1999. Causal Effects in Nonexperimental Studies : Reevaluation of the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94 , 1053 - 1062.
- [3] Abadie, A., and C.W. Imbens , 2006. Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica* , 74 , 235 - 267.
- [4] Rosenbaum ,P. R.,and D.B.Rubin , 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrics* , 70, 41 - 55.
- [5] Rosenbaum ,P.R.and D. B.Rubin, 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician*, 39 , 33 - 38.
- [6] Abadie,A., D. Drukker ,J. L. Herr , and G.W.Imbens , 2004. Implementing Matching Estimators for Average Treatment Effects in Stata. *The Stata Journal*, 4 , 290 - 311.

- [7] Angrist, J. D., and J. Hahn, 1999. When to Control for Covariates? Panel-Asymptotic Results for Estimates of Treatment Effects. *NBER Technical Working Paper* 241.
- [8] Heckman, J., H. Ichimura, and P. Todd, 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies*, 64, 605-654.
- [9] Rubin, D., and N. Thomas, 1996. Matching Using Estimated Propensity Scores: Relating Theory to Practice. *Biometrics*, 52, 249-264.
- [10] Robins, J., and A. Rotnitzky, 1995. Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association* , 90 , 429, 122 - 129.
- [11] Horvitz, D., and D. Thompson, 1952. A Generalization of Sampling Without Replacement from a Finite Population. *Journal of the American Statistical Association* , 47, 663 - 685.