应用微观计量经济学

Applied Microeconometrics

江　艇

中国人民大学经济学院

Last updated: April 8, 2019

Time: 周一晚 18:00-21:15
Venue: 明法 0502
Office: 明主 952
E-mail: ting.jiang@ruc.edu.cn
Phone: 8250-0201
Grading: 期末闭卷考试

Lecture 4　匹配

4.1　匹配的工作原理

一个假想例子

id	treat	x	y
1	1	1	y_1
2	1	1	y_2
3	1	2	y_3
4	1	3	y_4
5	0	1	y_5
6	0	2	y_6
7	0	2	y_7
8	0	2	y_8

	treat	control
$x=1$	y_1, y_2	y_5
$x=2$	y_3	y_6, y_7, y_8
$x=3$	y_4	

以估计 ATT 为例。若

$$E(Y^0|D,X)=E(Y^0|X)$$

则

$$\tau_1(X)=E(Y|X,D=1)-E(Y|X,D=0)$$

$$\begin{aligned}\tau_1 =&E(Y^1-Y^0|D=1)\\ =&E\big\{E(Y^1-Y^0|X,D=1)|D=1\big\}\\ =&E\big\{E(Y|X,D=1)-E(Y|X,D=0)|D=1\big\}\\ =&E(Y|D=1)-E_X\big\{E(Y|X,D=0)|D=1\big\}\end{aligned}$$

相反，若

$$E(Y^0|D)=E(Y^0)$$

$$\begin{aligned}\tau_1 =&E(Y^1-Y^0|D=1)\\ =&E(Y^1|D=1)-E(Y^0|D=0)\\ =&E(Y|D=1)-E_X\big\{E(Y|X,D=0)|D=0\big\}\end{aligned}$$

$$
\begin{aligned}
\tau_1(x=1) \; &= \left(\frac{y_1+y_2}{2}\right) - y_5 \\
\tau_1(x=2) \; &= y_3 - \left(\frac{y_6+y_7+y_8}{3}\right) \\
\tau_1(x=3) &\text{ undefined}
\end{aligned}
$$

$$
\begin{aligned}
\tau_1 &= \tau_1(x=1)\times\frac{2}{3} + \tau_1(x=2)\times\frac{1}{3} \\
&= \left(\frac{y_1+y_2+y_3}{3}\right) - \left(y_5\times\frac{2}{3} + \left(\frac{y_6+y_7+y_8}{3}\right)\times\frac{1}{3}\right) \\
&= \frac{1}{3}\left((y_1-y_5)+(y_2-y_5)+\left(y_3-\frac{y_6+y_7+y_8}{3}\right)\right) \\
&= \frac{1}{n_T}\sum_{i\in T\cup C_T}(2D_i-1)w_i y_i
\end{aligned}
$$

类似地,

$$
\begin{aligned}
\tau_0 &= \tau_0(x=1)\times\frac{1}{4}+\tau_0(x=2)\times\frac{3}{4}\\
&= \left(\left(\frac{y_1+y_2}{2}\right)\times\frac{1}{4}+y_3\times\frac{3}{4}\right)-\left(\frac{y_5+y_6+y_7+y_8}{4}\right)\\
&= \frac{1}{4}\left[\left(\frac{y_1+y_2}{2}-y_5\right)+(y_3-y_6)+(y_3-y_7)+(y_3-y_8)\right]
\end{aligned}
$$

$$
\begin{aligned}
\tau &= \tau(x=1)\times\frac{3}{7}+\tau(x=2)\times\frac{4}{7}\\
&= \left(\left(\frac{y_1+y_2}{2}\right)\times\frac{3}{7}+y_3\times\frac{4}{7}\right)-\left(y_5\times\frac{3}{7}+\left(\frac{y_6+y_7+y_8}{3}\right)\times\frac{4}{7}\right)\\
&= \frac{1}{7}\left[(y_1-y_5)+(y_2-y_5)+\left(y_3-\frac{y_6+y_7+y_8}{3}\right)\right.\\
&\qquad\left.+\left(\frac{y_1+y_2}{2}-y_5\right)+(y_3-y_6)+(y_3-y_7)+(y_3-y_8)\right]
\end{aligned}
$$

匹配与 OLS 的比较

- OLS 是一种控制 X 的参数方法，例如

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_i + \varepsilon_i$$

隐含着 $\tau(x=1) = \tau(x=2) = \tau(x=3) = \beta_1$.

$$\begin{aligned}
\tau(x=1) &= E(y|D=1, x=1) - E(y|D=0, x=1) \\
&= (\beta_0 + \beta_1 + \beta_2) - (\beta_0 + \beta_2) \\
&= \beta_1
\end{aligned}$$

同理，$\tau(x=2) = \tau(x=3) = \beta_1$.

又例如

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_i + \beta_3 D_i x_i + \varepsilon_i$$

隐含着 $\tau(x) = \beta_1 + \beta_3 x$.

$$\begin{aligned}\tau(x) &= E(y|D=1,x) - E(y|D=0,x) \\ &= (\beta_0 + \beta_1 + \beta_2 x + \beta_3 x) - (\beta_0 + \beta_2 x) \\ &= \beta_1 + \beta_3 x\end{aligned}$$

- 匹配是一种控制 X 的非参数方法，无需假定同质处理效应 (homogeneous treatment effect) 或处理效应关于 X 的函数形式。
- 但是，即使 $\tau(x)$ 各不相同并且高度非线性，能否认为 OLS 一致地估计了 $\tau = \int \tau(x) dF_X(x)$ 呢？更重要的是，匹配的识别假设和 OLS 的识别假设是一样的。所以我们需要深入探究**在异质处理效应前提下**，OLS 何时产生偏误。
- 要保证处理效应的 OLS 估计一致，下面两个条件中必须至少满足一个：
  - 条件期望函数确实为线性
  - 处理组和控制组的 X 的分布相同

  否则，存在由模型误设引起的外推偏误。

示例 1. 评估职业培训项目的效果 (Lalonde, 1986, *AER*; Dehejia and Wahba, 1999, *JASA*, 2002, *REStat*).

- 评估美国 1970 年代中期的某个职业培训项目。该项目是一项实地实验，招募一批劳动力市场上的弱势群体（戒毒瘾者、有犯案前科者、辍学者等），将其分为处理组和控制组，处理组个体可以获得 9-18 个月的工作机会。74、75 年为干预前，78 年为干预后。
- Lalonde (1986) 最先使用了这组数据，他先用处理组和控制组数据估计了这一职业培训项目的效果，然后用 PSID（收入动态面板调查）中的抽样数据代替控制组数据再次进行了估计，说明两者的差异。
- Dehejia and Wahba (1999, 2002) 发现，如果采用倾向得分匹配方法，那么即使用调查数据而非实际的控制组数据作为控制组，也能得到合意的结果。他们使用的数据与 Lalonde (1986) 有细微差别。
- 实验组的干预前结果和调查数据控制组的干预前结果有很大差异。

- 从另一个角度来理解：不同观测值对 OLS 估计的贡献如何？以下面的模型为例，

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 x_i + \beta_3 D_i x_i + \varepsilon_i$$

$$\begin{aligned}
\hat{\tau}_1 =& \bar{y}_T - \left(\hat{\beta}_0 + \hat{\beta}_2 \bar{x}_T\right) \\
=& \bar{y}_T - \left(\bar{y}_C + \hat{\beta}_2 \left(\bar{x}_T - \bar{x}_C\right)\right) \\
=& \bar{y}_T - \left(\bar{y}_C + \frac{\sum_{c\in C}\left(x_c - \bar{x}_C\right)\left(y_c - \bar{y}_C\right)}{\sum_{c\in C}\left(x_c - \bar{x}_C\right)^2} \cdot \left(\bar{x}_T - \bar{x}_C\right)\right) \\
=& \bar{y}_T - \frac{1}{n_C}\sum_{c\in C}\left(1 - \frac{\bar{x}_C - \bar{x}_T}{\overline{x_C^2} - \left(\bar{x}_C\right)^2} \cdot \left(x_c - \bar{x}_C\right)\right) \cdot y_C
\end{aligned}$$

$$\omega_i = 2.8084 - 0.0949 \cdot x_c$$

`re75` $= 156.7$ 的个体被赋予最高负权重 -12.05，但该个体的 outcome 是多少对于 `re75` $\in [0, 25.14]$ 的处理组个体而言是没有意义的。

匹配就是良好的控制

示例 2. 私立学校的经济回报 (Dale and Krueger, 2002, *QJE*).

The college matching matrix

		Private			Public			
Applicant group	Student	Ivy	Leafy	Smart	All State	Tall State	Altered State	1996 earnings
A	1		Reject	Admit		Admit		110,000
	2		Reject	Admit		Admit		100,000
	3		Reject	Admit		Admit		110,000
B	4	Admit			Admit		Admit	60,000
	5	Admit			Admit		Admit	30,000
C	6		Admit					115,000
	7		Admit					75,000
D	8	Reject			Admit	Admit		90,000
	9	Reject			Admit	Admit		60,000

Note: Enrollment decisions are highlighted in gray.

- 比较各种估计方法。
  - $\tau(A)=-5,\ \tau(B)=30.$
  - $\tau=(-5)\times(3/5)+30\times(2/5)=9.$
  - $\tau_1=(-5)\times(2/3)+30\times(1/3)=20/3.$
  - $\tau_0=(-5)\times(1/2)+30\times(1/2)=25/2.$
  - 全样本的组间均值差异：$\tau_{MD}=19.5.$
  - 匹配样本的组间均值差异：$\tau_{MDM}=20.$
  - 全样本的控制回归：$\tau_R=10.$
  - 匹配样本的控制回归：$\tau_{RM}=10.$
  - 饱和模型。
  - 匹配样本的加权简单回归。

- 控制回归是一种特殊的匹配估计，$\tau(X)$ 权重最高的是处理状态方差最大的 X 组。

Proof.

$$
\begin{aligned}
\tau_R =& \frac{Cov(Y, \tilde{D})}{Var(\tilde{D})} \\
=& \frac{E\left[(D - E(D|X))Y\right]}{E\left[D - E(D|X)\right]^2} \\
=& \frac{E\left[(D - E(D|X))E(Y|D, X)\right]}{E\left[D - E(D|X)\right]^2} \\
=& \frac{E\left\{(D - E(D|X))\left[E(Y|D = 0, X) + D\tau(X)\right]\right\}}{E\left[D - E(D|X)\right]^2} \\
=& \frac{E\left[(D - E(D|X))^2\tau(X)\right]}{E\left[D - E(D|X)\right]^2} \\
=& \frac{E\left[Var(D|X)\tau(X)\right]}{E\left[Var(D|X)\right]}
\end{aligned}
$$

$$Var(D|X) = P(D=1|X)(1-P(D=1|X))$$

$$\tau_R = \frac{\sum_x \tau(X=x)\left[P(D=1|X=x)(1-P(D=1|X=x))\right]P(X=x)}{\sum_x \left[P(D=1|X=x)(1-P(D=1|X=x))\right]P(X=x)}$$

在上例中,

$$\tau_R = \frac{\tau_A \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{3}{5} + \tau_B \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{5}}{\frac{2}{3} \cdot \frac{1}{3} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{5}} = \frac{-\frac{2}{3}+3}{\frac{2}{15}+\frac{1}{10}} = 10$$

而在匹配估计中，$\tau(X)$ 权重最高的是进入处理组概率最高的 X 组（τ_1）或进入控制组概率最高的 X 组（τ_0）。

$$\begin{aligned}\tau_1 &= \sum_x \tau(X=x)P(X=x|D=1) \\ &= \frac{\sum_x \tau(X=x)P(D=1|X=x)P(X=x)}{\sum_x P(D=1|X=x)P(X=x)}\end{aligned}$$

当处理状态独立于 X 时，OLS 估计和匹配估计等价。

- 验证 τ_R 的表达式。回到假想例子，假定采用 OLS 估计,

$$y_i = \beta D_i + \mu_1 \mathbf{1}(x_i = 1) + \mu_2 \mathbf{1}(x_i = 2) + \mu_3 \mathbf{1}(x_i = 3) + \varepsilon_i$$

$$\begin{aligned}
\sum D_i \varepsilon_i &= 0 \;\Rightarrow\; 4\beta + 2\mu_1 + \mu_2 + \mu_3 &&= y_1 + y_2 + y_3 + y_4 \\
\sum \mathbf{1}(x_i = 1)\varepsilon_i &= 0 \;\Rightarrow\; 2\beta + 3\mu_1 &&= y_1 + y_2 + y_5 \\
\sum \mathbf{1}(x_i = 2)\varepsilon_i &= 0 \;\Rightarrow\; \beta + 4\mu_2 &&= y_3 + y_6 + y_7 + y_8 \\
\sum \mathbf{1}(x_i = 3)\varepsilon_i &= 0 \;\Rightarrow\; \beta + \mu_3 &&= y_4
\end{aligned}$$

$$\hat{\beta} = \frac{1}{17}(4y_1 + 4y_2 + 9y_3 - 8y_5 - 3y_6 - 3y_7 - 3y_8)$$

根据前文公式,

$$\begin{aligned}
\tau_R &= \frac{\tau(X=1) \times \frac{2}{9} \times \frac{3}{7} + \tau(X=2) \times \frac{3}{16} \times \frac{4}{7}}{\frac{2}{9} \times \frac{3}{7} + \frac{3}{16} \times \frac{4}{7}} \\
&= \left(\left(\frac{y_1 + y_2}{2}\right) - y_5\right) \times \frac{8}{17} + \left(y_3 - \left(\frac{y_6 + y_7 + y_8}{3}\right)\right) \times \frac{9}{17}
\end{aligned}$$

- 进一步地，考虑饱和模型，

$$
\begin{aligned}
y_i = \beta_1 \cdot D_i \times \mathbf{1}(x_i = 1) + \beta_2 \cdot D_i \times \mathbf{1}(x_i = 2) + \beta_3 \cdot D_i \times \mathbf{1}(x_i = 3) \\
+ \mu_1 \mathbf{1}(x_i = 1) + \mu_2 \mathbf{1}(x_i = 2) + \mu_3 \mathbf{1}(x_i = 3) + \varepsilon_i
\end{aligned}
$$

这一模型等价于匹配估计，因为

$$
\begin{aligned}
\beta_k &= \tau(X = k) \\
\tau_1 &= \frac{1}{n_T} \sum_{i \in T} [\beta_1 \cdot \mathbf{1}(x_i = 1) + \beta_2 \cdot \mathbf{1}(x_i = 2) + \beta_3 \cdot \mathbf{1}(x_i = 3)] \\
\tau_0 &= \frac{1}{n_C} \sum_{i \in C} [\beta_1 \cdot \mathbf{1}(x_i = 1) + \beta_2 \cdot \mathbf{1}(x_i = 2) + \beta_3 \cdot \mathbf{1}(x_i = 3)] \\
\tau &= \frac{1}{n} \sum_{i=1}^{n} [\beta_1 \cdot \mathbf{1}(x_i = 1) + \beta_2 \cdot \mathbf{1}(x_i = 2) + \beta_3 \cdot \mathbf{1}(x_i = 3)]
\end{aligned}
$$

这种等价性只有当 X 离散时才能实现；当 X 连续时，模糊匹配 (fuzzy matching) 和基于函数形式的外推 (functional extrapolation) 会得到不同的结果。

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.135	.095	.086	.007	.003	.013
	(.055)	(.052)	(.034)	(.038)	(.039)	(.025)
Own SAT score ÷ 100		.048	.016		.033	.001
		(.009)	(.007)		(.007)	(.007)
Log parental income			.219			.190
			(.022)			(.023)
Female			−.403			−.395
			(.018)			(.021)
Black			.005			−.040
			(.041)			(.042)
Hispanic			.062			.032
			(.072)			(.070)
Asian			.170			.145
			(.074)			(.068)
Other/missing race			−.074			−.079
			(.157)			(.156)
High school top 10%			.095			.082
			(.027)			(.028)
High school rank missing			.019			.015
			(.033)			(.037)
Athlete			.123			.115
			(.025)			(.027)
Selectivity-group dummies	No	No	No	Yes	Yes	Yes

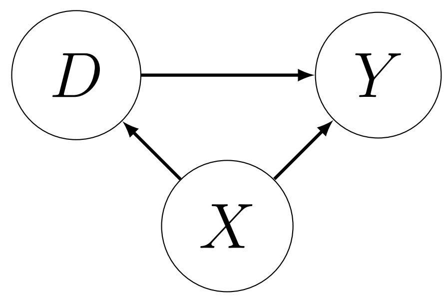
	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.212	.152	.139	.034	.031	.037
	(.060)	(.057)	(.043)	(.062)	(.062)	(.039)
Own SAT score ÷ 100		.051	.024		.036	.009
		(.008)	(.006)		(.006)	(.006)
Log parental income			.181			.159
			(.026)			(.025)
Female			−.398			−.396
			(.012)			(.014)
Black			−.003			−.037
			(.031)			(.035)
Hispanic			.027			.001
			(.052)			(.054)
Asian			.189			.155
			(.035)			(.037)
Other/missing race			−.166			−.189
			(.118)			(.117)
High school top 10%			.067			.064
			(.020)			(.020)
High school rank missing			.003			−.008
			(.025)			(.023)
Athlete			.107			.092
			(.027)			(.024)
Average SAT score of schools applied to ÷ 100				.110	.082	.077
				(.024)	(.022)	(.012)
Sent two applications				.071	.062	.058
				(.013)	(.011)	(.010)
Sent three applications				.093	.079	.066
				(.021)	(.019)	(.017)
Sent four or more applications				.139	.127	.098
				(.024)	(.023)	(.020)

4.2 匹配的实现细节

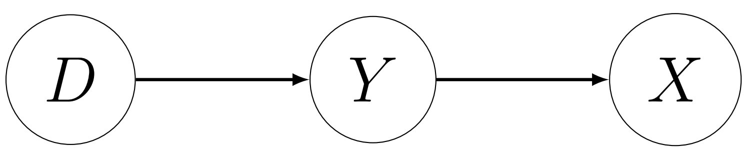
- 关键文献：Imbens (2015, *JHR*).

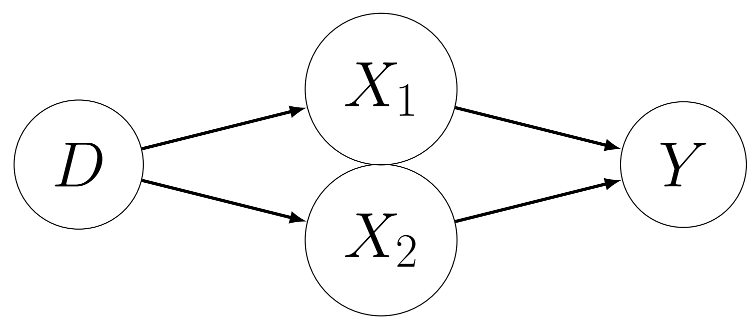
准备工作第 1 步：选择 X

- 基本原则："pre but not post".
- 必须控制

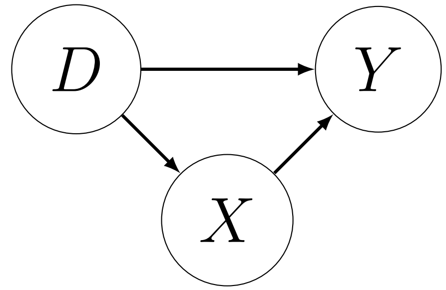


- 必须不能控制

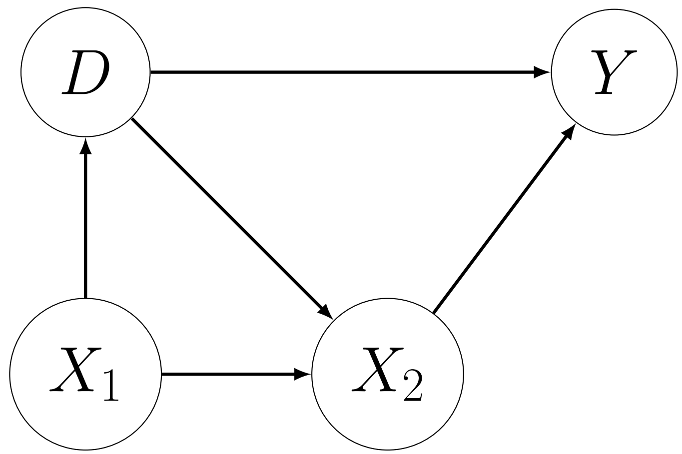




- 只有当关心直接效应时才控制



- 例子：一个两难选择



X_1: 不可观测的能力

D: 参加小学奥数竞赛

X_2: 受教育程度

Y: 收入水平

准备工作第 2 步：评估 overlap

- 计算 normalized difference (Imbens, 2015, *JHR*).

$$\Delta = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\left(s_C^2 + s_T^2\right)/2}}$$

$$s_C^2 = \frac{1}{N_C - 1} \sum_{c \in C} \left(X_c - \bar{X}_C\right)^2, \; s_T^2 = \frac{1}{N_T - 1} \sum_{t \in T} \left(X_t - \bar{X}_T\right)^2$$

- 如果 normalized difference 差异很大，则要考虑删截样本。

准备工作第 3 步：估计倾向得分

- $h(\cdot)$ 形式的选择不是为了提供因果解释，而只是为了更好地近似条件期望函数。

$$P(D=1|X) \triangleq \pi(X) = \frac{\exp(h(X)'\gamma)}{1+\exp(h(X)'\gamma)}$$

- 基于逐步回归的方程设定搜索：
  1. 先根据经验决定有哪些变量必须加入（如果没有先验信息，就只加入截距项）。
  2. 然后逐一加入其余所有一次项，对新增项的系数显著性进行 likelihood ratio test，统计量数值最大的一次项加入 $h(\cdot)$.
  3. 对余下的一次项重复步骤 2，直到本轮新增项系数的检验统计量最大值低于临界值 $C_{lin}=1$.
  4. 然后逐一加入所有二次项（包括平方项和交互项），进行类似于步骤 2-3 的操作，统计量临界值 $C_{qua}=2.71$.
- 根据最终确定的 $h(\cdot)$ 估计倾向得分。

准备工作第 4 步：删截样本

- 保留 $\alpha \le \pi(X) \le 1-\alpha$ 的样本 (Crump *et al.*, 2009, *Biometrika*)，其中 α 满足
$$\frac{1}{\alpha(1-\alpha)} = 2 \cdot E\left[\frac{1}{\pi(X)(1-\pi(X))} \middle| \frac{1}{\pi(X)(1-\pi(X))} \le \frac{1}{\alpha(1-\alpha)}\right]$$
- 经验法则：$\alpha = 0.1$.

匹配方法的大致分类

- 匹配对象：协变量匹配 (covariate matching) vs. 倾向得分匹配 (propensity score matching, PSM)
- 匹配精度：准确匹配 vs. 近似匹配。为了解决维度问题，常用 Mahalanobis distance (quadratic distance).

$$(X_c - X_t)'\Omega(X_t)^{-1}(X_c - X_t)$$

  其中 $\Omega(\cdot)$ 是样本协方差。
- 是否放回 (replacement)：有放回的匹配 vs. 无放回的匹配（bias 和 variance 的权衡）

- 加权方式：以估计 ATT 为例

$$\hat{\tau}_1 = \frac{1}{n_T} \sum_{t \in T} \left(Y_t - \hat{E}(Y|X_t, D = 0) \right)$$

关键在于得到 $\hat{E}(Y|X_t, D = 0)$.

– 最近邻 (nearest neighbor) k 匹配

$$\hat{E}(Y|X_t, D = 0) = \frac{1}{k} \sum_{c \in C_{t,k}} Y_c,\ C_{t,k} = \{j_1(t), \ldots, j_k(t)\}$$

其中 $j_m(t)$ 满足

$$\sum_{l \in C} \mathbf{1} \left(\|X_l - X_t\| \leq \|X_{j_m(t)} - X_t\| \right) = m,\ m = 1, \ldots, k$$

– 半径 (radius)/卡钳 (caliper) 匹配

$$\hat{E}(Y|X_t, D = 0) = \sum_{c \in C} \frac{\mathbf{1}(|X_c - X_t| \leq r) Y_c}{\mathbf{1}(|X_c - X_t| \leq r)}$$

– 核 (kernel) 匹配

$$\hat{E}(Y|X_t, D=0) = \sum_{c\in C} \frac{K\left(\frac{X_c - X_t}{h}\right) Y_c}{K\left(\frac{X_c - X_t}{h}\right)}$$

其中 $K(\cdot)$ 是关于原点对称的有界连续函数，且满足 $\int K(u)du = 1$.
核函数的选择：

▷ Uniform (rectangular) kernel:

$$K(u) = \frac{1}{2}\mathbf{1}\left(|u| < 1\right)$$

▷ Triangular kernel:

$$K(u) = (1 - |u|)\mathbf{1}\left(|u| < 1\right)$$

▷ Gaussian kernel:

$$K(u) = \phi(u)$$

▷ Epanechnikov kernel:

$$K(u) = 0.75(1 - u^2)\mathbf{1}\left(|u| < 1\right)$$

- 匹配的重点在于思想，而非具体实现手段。

示例 3. 信息不对称与融资决策 (Derrien and Kecskes, 2013, *JF*)

these restrictions on both groups of firms. We require that candidate control firms have the same two-digit SIC code as our treatment firms. We also require candidate control firms to be in the same total assets quintile, Q quintile, and cash flow quintile as our treatment firms.[6] We then retain candidate control firms that have the smallest difference in number of analysts compared to the corresponding treatment firms. We break any remaining ties based on the smallest differences in total assets, Q, and cash flow. To this end, we compute the difference between treatment firms and control firms for each of total assets, Q, and cash flow. We compute the rank of the difference for each of these three variables, and we compute the total rank across all three variables. We retain candidate control firms that have the lowest total rank.

基于协变量的最近邻匹配

- 固定的卡钳宽度或固定的最近邻数量。
- 不论是协变量匹配还是倾向得分匹配，可能都有必要预先进行一步准确匹配，例如性别、行业等。
- 不能用自抽样方差作为渐进方差估计 (Abadie and Imbens (2008, *ECMA*))。
- Abadie and Imbens (2006, *ECMA*) 证明了 NNM 估计量的一致性，并给出了正确的标准误公式。

- Abadie and Imbens (2011, *JBES*) 对 NNM 估计量进行了偏误修正（因为匹配不是精确的），并且说明 Abadie and Imbens (2006) 的标准误公式仍然正确。以 ATE 为例，

$$\hat{\tau}^{NNM} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i^1 - \hat{Y}_i^0\right)$$

$$\hat{Y}_i^1 = D_i Y_i + (1 - D_i)\frac{1}{M}\sum_{t \in T_i}\left\{Y_t + \hat{\mu}_1(X_i) - \hat{\mu}_1(X_t)\right\}$$

$$\hat{Y}_i^0 = (1 - D_i) Y_i + D_i\frac{1}{M}\sum_{c \in C_i}\left\{Y_c + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_c)\right\}$$

  其中 $\hat{\mu}_d(X)$ 是 $\mu_d(X) = E(Y|X, D = d)$ 的 OLS 估计。

- `teffects nnmatch` 和 `nnmatch` 都可以实现，但标准误估计有差异（不明原因）。

倾向得分匹配

- PSM 的识别。Rosenbaum and Rubin (1983, *Biometrika*) 证明,

$$Y^d \perp D|X \Rightarrow Y^d \perp D|\pi(X),\ d=0,1$$

因此

$$\begin{aligned}
\tau =& E(Y^1 - Y^0) \\
=& E\left\{E(Y^1 - Y^0|\pi(X))\right\} \\
=& E\left\{E(Y^1|D=1,\pi(X)) - E(Y^0|D=0,\pi(X))\right\} \\
=& E_{\pi(X)}\{E(Y|D=1,\pi(X)) - E(Y|D=0,\pi(X))\}
\end{aligned}$$

Proof: 因为 D 是二元变量, 因此只需证明其期望性质, 若 $Y^d \perp D|X$, 则

$$E(D|Y^d,\pi(X)) = E(D|\pi(X))$$

$$\begin{aligned}
LHS &= E\left[E(D|Y^d,X,\pi(X))\,\middle|\,Y^d,\pi(X)\right] \\
&= E\left[E(D|Y^d,X)\,\middle|\,Y^d,\pi(X)\right] \\
&= E\left[E(D|X)\,\middle|\,Y^d,\pi(X)\right] \\
&= E\left[\pi(X)\,\middle|\,Y^d,\pi(X)\right] \\
&= \pi(X)
\end{aligned}$$

$$\begin{aligned}
RHS &= E\left[E(D|X,\pi(X))|\pi(X)\right] \\
&= E\left[E(D|X)|\pi(X)\right] \\
&= E(\pi(X)|\pi(X)) \\
&= \pi(X)
\end{aligned}$$

- 上述结果成立的关键在于：给定 $\pi(X)$，X 和 D 之间不再相关，准确地说，

$$D \perp X|\pi(X)$$

也就是说，给定 $\pi(X)$，处理组和控制组的 X 的分布是相同的；也即，在分布的意义上，要想 balance X，只要 balance $\pi(X)$ 就够了。

$$E(D|X, \pi(X)) = E(D|\pi(X)) = \pi(X)$$

- 这样我们就能直观地理解 Rosenbaum and Rubin (1983) 的结果：试想如下的回归方程

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \varepsilon_i$$

如果遗漏 X，则对 β_1 估计的渐进偏误等于 $\beta_2\delta$，其中 δ 是 X 对 D 作回归得到的系数。但若给定 $\pi(X)$，X 和 D 不再相关，则忽略 X 不再产生偏误。

- 比较一下：
  - 基于 X 的匹配：使处理组和控制组之间的 X 相同。
  - 随机试验：使 X 和 ε 的分布相同。
  - 基于 $\pi(X)$ 的匹配：使 X 的分布相同。
- Abadie and Imbens (2016, *ECMA*) 给出了 PSM 估计量的渐进分布。
- `teffects psmatch`
  - 估计标准误默认采用 Abadie and Imbens (2016) 提出的正确形式。
  - 不能进行无放回匹配、核匹配。
- `psmatch2`
  - 能够得到和 `teffects psmatch` 完全相同的处理效应估计，但标准误估计有差异（采用的是 Abadie and Imbens (2006) 的形式，但这种形式没有考虑到 $\pi(X)$ 是估计值而非真实值）。
  - 提供了详尽的匹配细节。