Propensity score matching (PSM)

Jerzy Mycielski

Uniwersytet Warszawski

January 2024

Evaluation studies

Estimation of the treatment effect

- Examples:
  - evalution of the efficiency of the activation programmes in reducing unemployment
  - evaluation of the effects of the innovation grants

Ideal social experiment

1. members of the experimental group and control group has the same distribution of unobserved characteristics
2. members of the experimental group and control group has the same distribution of observed characteristics
3. the same questionnaire was used in experimental and control group - characteristics and outcomes measured in the same way
4. both groups are in the same economic environment

Evaluation studies cont.

- Ideal evolution study is based on the random experiment: for such a study the treatment effect can be measured simply as the difference between the mean outcome variable in the treatment and control groups
- in such a case assumptions (1)-(4) are satisfied by definition
- Traditional econometrics analysis of problems related to evolution studies based on nonexperimantal data was concentrated on assumption (1) - e.g. Heckman model
- in practice it seems that more important are the problems related to assumptions (2), (3), (4).
- PSM make possible to reduce bias of the estimated treatment effect related to invalidity of the assumption (2).

Notation

- $i \in I_1$ experimental group
- $j \in I_0$ control group
- $D \in \{0, 1\}$
  - $D = 0$ unit is untreated
  - $D = 1$ unit is treated
- X - vector of characteristics of the unit
- $\Pr(X) = \Pr(D = 1|X)$ conditional probability to be a member of experimental group (propensity score)
- Treatment effect is measured with the outcome variable Y
  - Y_0 value of the outcome variable if untreated
  - Y_1 value of the outcome variable if treated

Treatment effect

- Treatment Effect for a unit with characteristics given by X

$$E(Y_1 - Y_0 | D = 1, \mathrm{X})$$

- Average Treatment Effect - ATE

$$ATE = E(Y_1 - Y_0)$$

- Average Treatment Effect on the treated - ATE_1

$$ATE_1 = E(Y_1 - Y_0 | D = 1)$$

Treatment effect cont.

- We observe Y

$$Y = DY_1 + (1 - D)\, Y_0$$

- Every unit has the observed effect and counter-factual effect

	I_0	I_1
$D = 0$	Y_0	Y_1
$D = 1$	Y_0	Y_1

- It is not possible to observe the outcome of the treatment for the control group $E(Y_0 | D = 1, \mathrm{X})$ (counter factual effect)

Intuition of PM

- Treatment effect is based on comparison of the outcomes for treated units and untreated units.
- The more similar is the untreated unit to the treated unit the larger is the weight of this observation in this comparison
- Estimate of the treatment effect for unit i

$$Y_{1i} - \sum_{j \in I_0} W_{N_0,N_1}(i,j)\, Y_{0j}$$

- Estimated of the average treatment effect (ATE)

$$\sum_{i \in I_1} w_{N_0,N_1(i)} \left[Y_{1i} - \sum_{j \in I_0} W_{N_0,N_1}(i,j)\, Y_{0j} \right]$$

- Various PM variants differ in the way the weight functions $w_{N_0,N_1(i)}$ and $W_{N_0,N_1}(i,j)$ are defined.

Statistical independence

- Definition

$$\Pr(D = 1 | Y_0, Y_1, X) = \Pr(D = 1 | X) \tag{1}$$

- Selection based only on observable variables
- We assume that there is no self selection based on unobserved variables
- Excluded is the selection based on unobserved variables correlated with (influencing) the outcome variable as in Hackman model
- Equivalent notation

$$(Y_0, Y_1) \perp D | X$$

- Stable Unit Treatment Assumption SUTVA - outcome of one unit is independent from the outcomes of other units

Strong statistical independence

- Common support assumption

$$0 < \Pr(X) < 1 \tag{2}$$

- Strong statistical independence: (1) (2) satisfied
- Condition(2) is necessary condition for application of PSM.
- If $\Pr(X) = 0$ or $\Pr(X) = 1$, the such at unit is never treated or always treated and then forming the pair of the treated and untreated observations is impossible
- If condition (2) is invalid for some X we have to restrict our population to subpopulation for which this condition is satisfied
- Define

$$S_0 = Supp(X|D = 0)$$
$$S_1 = Supp(X|D = 1)$$

- Common support

$$S = S_0 \cap S_1$$

Bias decomposition in evaluation studies

- Difference in outcomes in treatment and control groups:

$$E(Y_0|D=1, X \in S_1) - E(Y_0|D=0, X \in S_0) = B_1 + B_2 + B_3$$

- Bias related to different in supports

$$\begin{aligned} B_1 = & E[E(Y_0|D=1, X \in S_1 \setminus (S_0 \cap S_1))|D=1] \\ & - E[E(Y_0|D=0, X \in S_0 \setminus (S_0 \cap S_1))|D=0] \end{aligned}$$

- Bias related to differences in distributions of observed characteristics in experimental and control groups:

$$\begin{aligned} B_2 = & E[E(Y_0|D=0, X \in S_0 \cap S_1)|D=1] \\ & - E[E(Y_0|D=0, X \in S_0 \cap S_1)|D=0] \end{aligned}$$

Bias decomposition in evaluation studies cont.

- Bias related to differences in distributions of unobserved characteristics in experimental and control groups:

$$B_3 = E\left[E\left(Y_0|D=1\right) - E\left(Y_0|D=0\right)|D=1, X \in S_0 \cap S_1\right]$$

- Application of PSM makes possible the correction of the bias B_2 and the determination of the common support for which bias B_1 is eliminated.

Implications of the statistical independence

- Strong statistical independence implies that, the average counter factual effect can be calculated as follows

$$\begin{aligned} E(Y_0|D=1) &= E[E(Y_0|D=1,\mathrm{X})|D=1] \\ &= E[E(Y_0|D=0,\mathrm{X})|D=1] \end{aligned}$$

- Assume that we succeeded to estimate (parametrically or non parametrically) the regression $\widehat{r}_0(\mathrm{X}_i) = E(Y_0|D=0,\mathrm{X}_i)$ and $\widehat{r}_1(\mathrm{X}_i) = E(Y_1|D=1,\mathrm{X}_i)$
- Estimate of ATE_1 can be obtained as the following mean

$$\frac{1}{N_1}\sum_{i\in I_1}\left(\widehat{r}_1(\mathrm{X}_i) - \widehat{r}_0(\mathrm{X}_i)\right)$$

- The assumption needed for this estimator is that $Y_0 \perp D|\mathrm{X}$
- In fact to estimate ATE_1 sufficient condition is that of independence of the conditional expectations from X

$$E(Y_0|D=1,\mathrm{X}) = E(Y_0|D=0,\mathrm{X})$$

Implications of the statistical independence cont.

- Strong statistical independence implies that (Resenbaum, Rubin 1983)

$$(Y_0, Y_1) \perp D | \Pr(X)$$

- Proof

$$\begin{aligned} E[D|Y_0, Y_1, \Pr(X)] &= E\{E[D|Y_0, Y_1, X]|\Pr(X)\} \\ &= E[\Pr(D=1|Y_0, Y_1, X)|\Pr(X)] \\ &= E\{\Pr(D=1|X)|\Pr(X)\} \\ &= E\{\Pr(X)|\Pr(X)\} = \Pr(X) \\ &= E[D|\Pr(X)] \end{aligned}$$

Implications of the statistical independence

- Resenbaum and Rubina result implies, that the independence of outcome from the membership in the group (experimental/control) if satisfied not only conditional on all the characteristics but also conditional on propensity score $\Pr(X)$ only
- Notice that if we have observations from experimental and control groups with identical $\Pr(X)$, then ATE_1 can be calculated as follows:

$$E[Y_1|D=1,\Pr(X)]-E[Y_0|D=0,\Pr(X)]=E[Y_1-Y_0|\Pr(X)]$$

- Choosing the observations to the control group we can base the choice on the similarity of the propensity score $\Pr(X)$ only

In PSM control group is chosen from the sample of untreated units on the basis of the similarity of the estimated values of the propensity scores $\Pr(X)$

PSM - estimation of the propensity score

- Parametric methods(probit, logit)
- Nonparametric methods
- Crucial element in estimation of $\Pr(X)$ is to make sure that the common support assumption is satisfied
  - range of the predicted values should be similar
  - predictors must not perfectly predict outcomes
- In some sens the model for$\Pr(X)$ should not be "too well fitted"

PSM - choosing the observations to the control group

- Selecting the observations wit replacement or without replacement
- Number of the untreated units chosen for one treated: 1 do 1, 1 do n
- Selection algorithm:
  - nearest neighbor
  - nearest neighbor with caliper
  - radius matching
  - kernel method

$$w_{ij} = \frac{K\left(\frac{\Pr(X_i)-\Pr(X_j)}{h}\right)}{\sum_{j \in I_0} \left(\frac{\Pr(X_i)-\Pr(X_j)}{h}\right)}$$

where h is the bandwidth

- The choice of the algorithm depends mainly on the sample size

PSM - checking the matching quality

- In the best case the control group selected on the basis of PSM should is large and has the distribution of observable characteristics identical to experimental group
- When choosing the selection algorithm we should take into account two factors:
  - the bias of evaluated treatment effect: the smaller the more similar are the units in control group to the units in treatment group
  - the variance of the estimate of the treatment effect: the smaller the larger is the control group
- The quality of matching is tested by comparing mean characteristics in treatment and control groups
- For this effect we can use the formal tests for equal means

PSM - advantages and disadvantages

- Advantages
  - In the case of PSM we make no assumptions about the functional form of the relationship between the expected outcomes and the values of characteristics - relationship $E(Y_1|D=1,X)$ i $E(Y_0|D=1,X)$ is left unspecified and can have quite general form
  - eliminates two most important sources of the estimation bias
- Disadvantages
  - the strong statistical independence assumptions must be satisfied
    - absence of self selection
    - absence of selection based on unobserved characteristics
  - needs large number of observations - especially the sample of untreated observations should be large to make possible selection of the untreated units sufficiently similar to treated observations