Economics of deforestation

Matching

Elías Cisneros

Matching

Literature

- Angrist, J. D. and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
  ISBN 0691120358
- Ho, D. E., K. Imai, G. King, and E. A. Stuart (2007): Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15 (3): 199–236

Intuition

- "matching amounts to covariate-specific treatment-control comparisons, weighted together to produce a single overall average treatment effect" (Angrist and Pischke, 2008)
- comparison of apples-with-apples vs. apples-with-peaches

Outcomes

$$\text{Potential outcomes } Y_i = \begin{cases} Y_{0i}, & \text{if } T = 0 \\ Y_{1i}, & \text{if} T = 1 \end{cases}$$

$$\text{Observed outcome: } Y_i = Y_{0i} + (Y_{1i} - Y_{0i})T_i$$

Effects

Average treatment effects on the treated (ATT):

$$ATT = E[Y_{1i} - Y_{0i}|T_i = 1] = E[Y_{1i}|T_i = 1] - \underbrace{E[Y_{0i}|T_i = 1]}_{\text{Counterfactual}}$$

Comparison of observed data:

$$\begin{aligned} E[Y_i|T_i = 1] - E[Y_i|T_i = 0] &= E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 0] \\ &= \underbrace{E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 1]}_{\text{ATT}} \\ &\quad + \underbrace{E[Y_{0i}|T_i = 1] - E[Y_{0i}|T_i = 0]}_{\text{Selection bias}} \end{aligned}$$

Non-Random research

& the conditional independence assumption

Conditional independence assumption (CIA) / Selection on observables / ignorability/ absence of omitted variable bias:

$$\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp T_i | X_i$$

- X_i must include all variables that are causally associated with T_i and effect Y_i
- X_i are truly pre-treatment & not consequence of T_i
- Independencd of units
  - no serial autocorrelation
  - no spatial autocorrelation
- treatment mode is equal to each unit

Non-Random research

& the conditional independence assumption

$$
\begin{aligned}
& E[Y_{1i}|X_i, T_i = 1] - E[Y_{0i}|X_i, T_i = 0] \\
& = \underbrace{E[Y_{1i}|X_i, T_i = 1] - E[Y_{0i}|X_i, T_i = 1]}_{\text{Average ATT at } X_i} + \underbrace{E[Y_{0i}|X_i, T_i = 1] - E[Y_{0i}|X_i, T_i = 0]}_{\text{Selection bias}} \\
& \qquad \text{as } \{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp T_i|X_i \text{ it follows:} \\
& = E[Y_{1i}|X_i, T_i = 1] - E[Y_{0i}|X_i, T_i = 1] \\
& = E[Y_{1i} - Y_{0i}|X_i, T_i = 1] \\
& = E[Y_{1i} - Y_{0i}|X_i]
\end{aligned}
$$

→ we can compare observed realizations when we condition on X_i

Non-Random research

& Effects

Overall or unconditional average effect:

$$\begin{aligned}
&E\{E[Y_{1i}|X_i, T_i = 1] - E[Y_{0i}|X_i, T_i = 0]\} \\
&= E\{E[Y_{1i} - Y_{0i}|X_i]\} \\
&= E[Y_{1i} - Y_{0i}]
\end{aligned}$$

Average causal effect of treated: :

$$\begin{aligned}
&E\{E[Y_{1i}|X_i, T_i = 1] - E[Y_{0i}|X_i, T_i = 0] \,|T_i = 1\} \\
&= E\{E[Y_{1i} - Y_{0i}|X_i] \,|T_i = 1\} \\
&= E[Y_{1i} - Y_{0i} \,|T_i = 1]
\end{aligned}$$

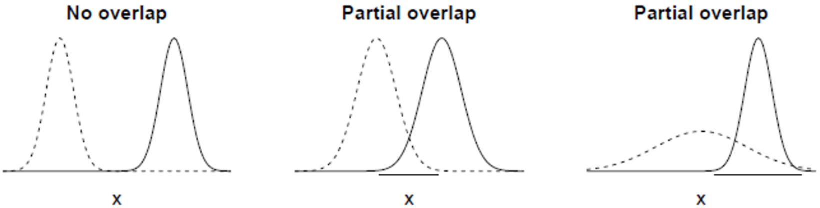
If causal effects are constant over i, then $ATE = ATT$

Non-Random research

& the Assumption on Overlap

Overlap / Common support:

$$0 < P(T_i = 1 | X_i) < 1 \tag{1}$$

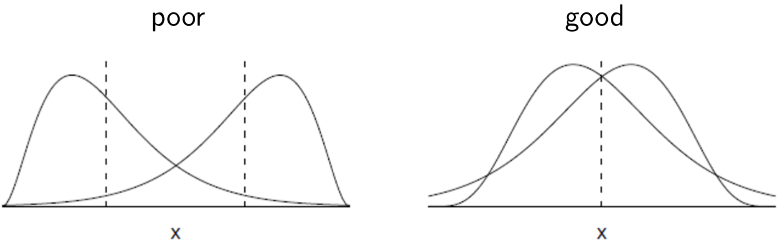


Lack of overlap = Regions in the covariate space for which there are no controls (or treatments)

Based on Gelman & Hill 2006

Challange in Matching:

Finding a balanced dataset



Imbalance = Different distributions of covariates in treatment and control groups

Based on Gelman & Hill 2006

Non-Random research

Computation of effects, e.g. ATT

$$E\{E[Y_{1i}|X_i, T_i = 1] - E[Y_{0i}|X_i, T_i = 0]\ |T_i = 1\}$$

Both, regression *and* matching, rely on the CIA!

- Regression:
  - $Y_i = \alpha + \beta T_i + X_i\gamma + \varepsilon_i$
  - weights by the variance of observations
- Matching (in the discrete case):

$$E[Y_{1i} - Y_{0i}| T_i = 1] = \sum_x \{\ E[Y_i|X_i, T_i = 1] - E[Y_i|X_i, T_i = 0]\ \}\ P(X_i = x|T_i)$$

  - $P(\cdot)$ is the probability mass function for X_i given $T_i = 1$
  - weigting covariate-specific estimates by the distribution of covariates among the treated

- **Individual matching:**
  - $\sum_{i=1}^{N}[Y_i(T = 1) - Y_{M(i)}(T = 0)]$
  - comparison of paired treated and matched controls

Exact matching (Example)

Nearest neighbour matching without replacement

id	T	Y	X_1	X_2
1	0	5	1	1
2	0	6	2	2
3	0	7	3	1
4	0	8	1	2
5	0	9	2	1
6	1	10	3	2
7	1	9	1	1
8	1	13	2	2
9	1	14	3	1
10	1	20	1	2

Treatment effect = 7.5

Curse of dimensionality

The curse of dimensionality:

- $X_1 \in \{1, ..., 10\}$ and $X_2 \in \{1, ...10\}$
- $10 \times 10 = 100$ possible combinations,
- etc.

- unlikely to find matches for all treated individuals

Search problem

- exponentially asymptotic

Solution: reduce dimensionality with Propensity score matching or Mahalanobis distance matching

Other matching methods

Criterion

- Exact matching
- Propensity score matching

$$p(x) = Pr(T = 1|X = x)$$

- Mahalanobis distance matching

$$md(X_i, X_j) = [X_i - X_j)^T S^{-1}(X_i - X_j)]^{1/2}$$

Other matching methods

Techniques

- Nearest neighbour matching

  Each treatment unit is matched to the control unit with the closest weight (e.g. propensity score), matching to one (1:1) or more control units (1:m) is possible; with or without replacement
- Caliper / radius matching
  - Nearest Neighbor matching with tolerance levels. May drop control observations
- Kernel / local linear matching
  - Using a weighted average of all controls to construct counterfactual for each treated unit
- etc.

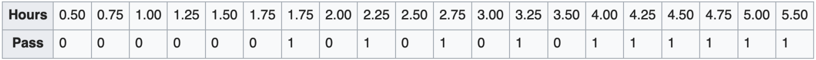
Propensity score matching analysis: Step by Step

1. Predict treatment probablity using a large number of predictors:
   $p(x) = Pr(T = 1|X = x)$
   Usually estimated using a logit or probit regression
2. Extract propensity scores (i.e. probabilities) from step 1
3. Match control and treatment units with minimum difference in propensity scores
4. Calculate TOT/ATE or regress matching covariates along with treatment indicator on outcome using the matched data set

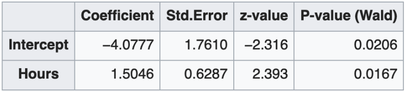
Propensity matching example

Logistic regression

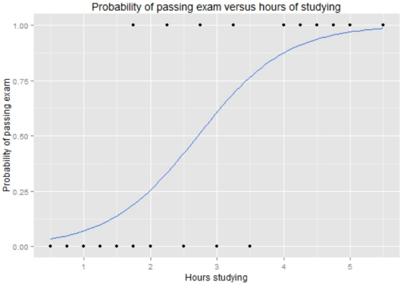
Data on exam outcomes versus hours of studying:



Hours	0.50	0.75	1.00	1.25	1.50	1.75	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	4.00	4.25	4.50	4.75	5.00	5.50
Pass	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1



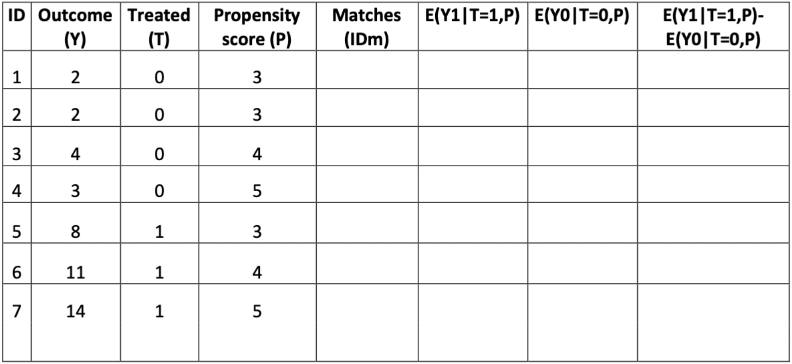
	Coefficient	**Std.Error**	**z-value**	**P-value (Wald)**
Intercept	−4.0777	1.7610	−2.316	0.0206
Hours	1.5046	0.6287	2.393	0.0167



$$\text{Probability of passing exam} = \frac{1}{1+exp(-(1.5046\cdot\text{Hours}-4.0777))}$$

Image source: `https://en.wikipedia.org/wiki/Logistic_regression`

Doing a nearest neighbor propensity score matching by hand



ID	Outcome (Y)	Treated (T)	Propensity score (P)	Matches (IDm)	E(Y1|T=1,P)	E(Y0|T=0,P)	E(Y1|T=1,P)-E(Y0|T=0,P)
1	2	0	3				
2	2	0	3				
3	4	0	4				
4	3	0	5				
5	8	1	3				
6	11	1	4				
7	14	1	5				

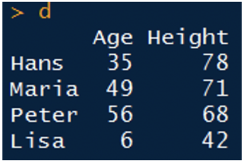
$$ATT_{PSM} = \frac{1}{N_T}[\sum_{i \in T} Y_i - \sum_{j \in C} w(i,j) Y_j^C]$$

With 1:1 matching $w(i,j) = 1$

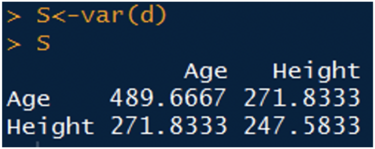
Mahalanobis distance

$$md(X_i, X_j) = [X_i - X_j)^T S^{-1}(X_i - X_j)]^{1/2}$$

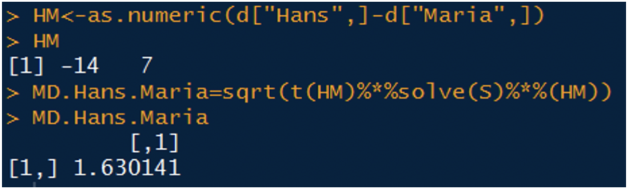
Covariate matrix: $S = cov[X_i, X_j] = E\{(X_i - E[x_i])(X_j - E[X_j])\}$



```
> d
      Age Height
Hans   35     78
Maria  49     71
Peter  56     68
Lisa    6     42
```



```
> S<-var(d)
> S
            Age   Height
Age    489.6667 271.8333
Height 271.8333 247.5833
```



```
> HM<-as.numeric(d["Hans",]-d["Maria",])
> HM
[1] -14   7
> MD.Hans.Maria=sqrt(t(HM)%*%solve(S)%*%(HM))
> MD.Hans.Maria
         [,1]
[1,] 1.630141
```