

Quantitative Economics for the Evaluation of the European Policy

Dipartimento di Economia e Management

Co-funded by the
Erasmus+ Programme
of the European Union



Project funded by
European Commission Erasmus + Programme –Jean Monnet Action
Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

Irene Brunetti Davide Fiaschi Angela Parenti¹

28/11/2016

¹ireneb@ec.unipi.it, davide.fiaschi@unipi.it, and aparenti@ec.unipi.it.

Introduction

We may want to estimate the effect of a policy, in situations when:

- controlled randomization is impossible;
- there are no convincing natural experiments providing a substitute to randomization (a RDD, a good instrument ...).

Regressions and matching can offer a way to estimate treatment effects, under the strong assumption of selection on observables (or unconfoundedness, or conditional independence):

- the selection into treatment is completely determined by variables that can be observed by the researcher;
- conditioning on these observable variables, the assignment to treatment is random.

Introduction

- Rosenbaum and Rubin (1983) proposed **propensity score matching** as a method to reduce the bias in the estimation of treatment effects with observational data sets.
- These methods have become increasingly popular in medical trials and in the evaluation of economic policy interventions.
- **Warning:** Matching **STILL** does not allow to control for selection bias that arises when the assignment to the treatment is done on the basis of non-observables.

Steps for designing observational studies

- As suggested by Rubin (2008), we have to design observational studies to approximate randomized trial in order to obtain objective causal inference.
- Are sample sizes in the data set adequate?
- Who are the decision makers for treatment assignment and what measurements were available to them?
 - What are the 'key' variables?
 - How the treatment conditions were assigned?
- Are key covariates measured well?
- Can balance be achieved on key covariates?

Observational studies

- To adjust any difference in average outcomes for differences in pre-treatment characteristics (not being affected by the treatment):
 - **Model-based imputation methods** (e.g., regression models);
 - Matching methods;
 - **Methods based on propensity score**;
 - Stratification;
 - Weighting or Mixed methods.

Matching methods based on propensity score

- Matching methods are like completely randomized experiments except that the probabilities of treatment assignment are allowed to depend on covariates, and so can vary from unit to unit.
- Two conditions:
 - **Unconfoundedness**: assignment to treatment is independent of the outcomes, conditional on the covariates:

$$Y(0); Y(1) \perp D|X \quad (1)$$

- **Overlap** (or common support condition): the probability of assignment is bounded away from zero and one:

$$0 < Pr(D = 1|X) < 1 \quad (2)$$

- The assignment probabilities, p_i , are called **propensity scores**.

Unconfoundedness and Overlap

Unconfoundedness

- The reduction to a paired-comparison should only be applied if unconfoundedness is a plausible assumption based on the data and a detailed understanding of the institutional set-up by which selection into treatment takes place.

Overlap

$$0 < Pr(D = 1|X) < 1$$

- The assignment mechanism can be interpreted as if, within subpopulations of units with the same value for the covariate, a completely randomized experiment was carried out.
- We can analyze data from subsamples with the same value of the covariates, as if they came from a completely randomized experiment.

Unconfoundedness and Overlap

- In their seminal article, Rosenbaum and Rubin (1983) define the treatment to be **strongly ignorable** when both unconfoundedness and overlap are valid.
- Given these two key assumptions of **unconfoundedness and overlap** one can identify the average treatment effects.

ATE identification under unconfoundedness

- Given unconfoundedness, the following equality holds:

$$E[Y(D)|X = x] = E[Y(d)|D = d; X = x] = E[Y|D = d; X = x]$$

- Thus one can estimate ATE by first estimating the average treatment effect for a sub-population with covariates $X = x$:

$$E[Y(1) - Y(0)|X = x] = E[Y(1)|X = x] - E[Y(0)|X = x] = \\ E[Y|X; D = 1] - E[Y|X; D = 0]$$

- We need to estimate $E[Y(d)|D = d; X = x]$ for all values of D and x in the support of these variables.
- If the overlap assumption is violated at $X = x$, it would be infeasible to estimate $E[Y(1)|X; D = 1] - E[Y(0)|X; D = 0]$.

Linear regression

- We could just run the following regression:

$$Y_i = \alpha + \tau D_i + \sum_k \beta_k X_{ki} + \epsilon_i \quad (3)$$

- Example: Y is an health outcome, D is whether an individual smokes and X_k are all the variables that we think affect the probability of smoking.
- This estimation is valid only if the probability of smoking is just a linear function of X_k .
- If the relationship is non linear, estimations may be very sensitive to specifications.

Example proposed by Grilli-Rampichini: Effect of participation in a job training program on individuals earnings

Data used by Lalonde (1986)

- Interest in the possible effect of participation in a job training program on individuals earnings in 1978
- This dataset has been used by many authors (Abadie et al.2004, Becker and Ichino, 2002, Dehejia and Wahba, 1999).
- They use of a subset of the data constructed by Dehejia and Wahba (1999, see their paper for details).
- **Treatment variable:** participation in the job training program
- **Outcome variable:** re78: 1978 earnings of the individuals in the sample in terms of 1978 dollars.

Example: covariates

To identify similar individuals they use following **observable pre-treatment covariates**:

- age,
- years of education,
- real yearly earnings in 1974,
- real yearly earnings in 1975,
- afro-american,
- hispanic-american,
- married,
- more than grade school but less than high school education,
- unemployed in 1974,
- unemployed in 1975.

Regression-based estimation under unconfoundedness

We need to adjust any difference in average outcomes for differences in pre-treatment characteristics (not being affected by the treatment).

- We can adjust via specification of a conditional model for the potential outcome \implies regression models.
- In a standard regression approach, unconfoundedness is implicitly assumed together with other functional or distributional assumptions.

$$Y_i^{obs} = \alpha + \tau D_i + \beta X_i + \epsilon_i \quad (4)$$

with usual exogeneity assumption that $\epsilon_i \perp D_i, X_i$.

- the slope τ of the treatment indicator is an estimator of the average treatment effect.

Regression: which problems?

To identify causal effects, unconfoundedness is not enough, to achieve ignorability, we need also overlap.

- If the difference between the average values of the covariates in the two groups is large, the results are sensitive to the linearity assumption
- More generally, since we do not know the exact nature of dependence of the assignment on the covariates, this results in increased sensitivity to model and to a-priori assumptions.
- Choice of covariates to be included in the model strongly affects results (specification of propensity score).
- In order to avoid model dependence we can consider matching techniques: **exact matching or propensity score matching**.

Matching

- Under unconfoundedness, the basic idea is to find in a large group of non-treated units, units similar to the treated subjects in all relevant pre-treatment characteristics X .
- Since conditioning on all relevant covariates is limited in the case of a high dimensional vector X , we will use propensity score, i.e. the probability of being treated given observed characteristics X .

Matching vs OLS

The main assumption underlying the matching approaches (unconfoundedness) is the same as OLS \implies as OLS, the matching is as good as its X are!

Why matching could be better than OLS?

- The additional common support condition focuses on comparison of comparable subjects.
- matching is a non-parametric technique: it avoids potential misspecification of $E(Y(0)|X)$.

Matching and regression: the dimensionality problem

- Both exact matching and regression may not be feasible if the sample is small, the set of covariates is large and many of them are multivalued, or, worse, continue.
- If the number of cells is very large with respect to the size of the sample it is possible that cells contain only treated or only control subjects.
- Conditioning on all relevant covariates is limited in the case of a high dimensional vector X .

The propensity scores

- Rosenbaum and Rubin (1983) suggest the use of a balancing score. One possible balancing score is the **propensity score**, i.e. the probability to be treated given observed characteristics X :

$$e(X) = Pr(D = 1|X = x) = E[D|X = x]$$

- The propensity score is a balancing score because:

$$Pr(D_i = 1|X_i; e(X_i)) = Pr(D_i = 1|X_i) = e(X_i)$$

- If treatment assignment is strongly ignorable given X , then it is strongly ignorable given any balancing score, i.e. D_i is independent of X_i given the propensity score.

The propensity scores

- At any value of a balancing score, the difference between the treatment and control means is an unbiased estimate of the average treatment effect at that value of the balancing score if treatment assignment is strongly ignorable.
- Given the strongly ignorable treatment assignment and the covariance adjustment on a balancing score, we obtain an unbiased estimates of treatment effects.

The role of propensity score

Many of the procedures for estimating and assessing causal effects under unconfoundedness involve the propensity score.

- If the balancing hypothesis is satisfied, observations with the same propensity score must have the same distribution of observable (and unobservable) characteristics independently of treatment status.
- For a given propensity score, exposure to treatment is random and therefore treated and control units should be on average observationally identical.
- Usually, given a set of pre-treatment variables, unconfoundedness is viewed as a reasonable approximation to the actual assignment mechanism, with only vague a priori information about the dependence of the propensity score on the pre-treatment variables.
- The true propensity score is generally unknown, so that the propensity score needs to be estimated non-parametrically.

The role of propensity score

Once estimated the propensity score we can estimate the average effect of treatment given the propensity score.

- Ideally in these steps, we would like to:
- match treatment and controls with exactly the same (estimated) propensity score;
- compute the effect of treatment for each value of the (estimated) propensity score;
- obtain the average of these conditional effects.
- This is infeasible in practice because it is rare to find two units with exactly the same propensity score.

Propensity score matching

- There are several alternative procedures to estimate the ATET given the propensity score, including:
 - Nearest neighbor matching on the score.
 - Radius matching on the score.
 - Stratification or Interval Matching.
 - Kernel matching on the score.
- They vary in: the method used to select the matches, the weight associated with each match.
- There is a trade-off between quality and quantity of matches.

Propensity score matching: Common support

- $0 < Pr(D = 1|X) < 1 \implies$ If the probability of treatment given X is equal to one, there is no observation with X among untreated.
- Counterfactuals for the treated cannot be evaluated on this point and vice-versa.
- We have to ensure that there is common support in the data.

Matching in practice

In practice:

- estimate $e(X)$ with a flexible method
- Trace the distribution of scores in the two populations $D = 1$ and $D = 0$.
- Determine the support of e that is common to the two populations
- Get rid of observations with e out of the common support
- Estimate ATET on the common support with some (or several) methods.
- Compute variances for ATET estimators.

Propensity Score Estimation

- The dimensionality problem is moved to the estimation of the propensity score: how to choose the covariates to be included in the propensity score?
- Standard probability models can be used to estimate the propensity score, e.g. a logit (or a probit) model:

$$Pr(D_i = 1|X_i) = \frac{\exp(h(X_i))}{1 + \exp(h(X_i))} \quad (5)$$

where $h(X_i)$ is a function of covariates with linear and higher order terms.

Propensity Score Estimation

The inclusion of higher order terms in $h(X_i)$ is determined only by the need to obtain an estimate of the propensity score that satisfies the balancing property.

- The specification of $h(X_i)$ that satisfies the balancing property is usually more parsimonious than the full set of interactions needed to match cases and controls on the basis of observables,
- the propensity score reduces the dimensionality problem of matching treated and control units on the basis of the multidimensional vector X .

An algorithm for the estimation of the propensity score

Procedure implemented by Dehejia & Wahba 1999 and Becker & Ichino 2002.

- Start with a parsimonious logit or probit function to estimate the score.
- Stratify the sample over small propensity scores intervals
- For each covariate, test whether the means for the treated and the controls are not statistically different for each interval.
- If it is not the case for some covariate, improve the specification (including more interactions or higher order terms) and start again.

Checking for common support (from the Handbook of Impact Evaluation, WB 2010)

- Empirically, trim extreme values to ensure common support, or drop control observations that have a lower propensity score than the minimum propensity score of the treated observations and treated observations that have higher propensity score than the maximum propensity score of control observations.
- Problem if the common support is very limited: the estimated effect may be different from ATET or ATE, since it will be estimated on a very specific subset of the population.

Nearest and radius matching with replacement

The Nearest Neighbor matching:

- NN match treated and control units taking each treated unit and searching for the control unit with the closest propensity score; i.e., the Nearest Neighbor.
- Although it is not necessary, the method is usually applied with replacement, in the sense that a control unit can be a best match for more than one treated unit.
- Once each treated unit is matched with a control unit, the difference between the outcome of the treated units and the outcome of the matched control units is computed.
- The ATET of interest is then obtained by averaging these differences.
- All treated units find a match.

Radius matching

- Each treated unit is matched only with the control units whose propensity score falls into a predefined neighborhood of the propensity score of the treated unit.
- Drop the unmatched controlled units.

Formally, denote by $C(i)$ the set of control units matched to the treated unit i with an estimated value of the propensity score, $e_i(x_i) = p_i$:

- **Nearest neighbor matching** sets $C(i) = \min\|p_i - p_j\|$ which is a singleton set unless there are multiple nearest neighbors.
- **In radius matching**, $C(i) = \{p_j \mid \|p_i - p_j\| < r\}$, i.e. all the control units with estimated propensity scores falling within a radius r from p_i are matched to the treated unit i .

Matching estimator

Denote the number of controls matched with observation $i \in T$ by N_i^C and define the weights $w_{ij} = \frac{1}{N_i^C}$ if $j \in C(i)$ and $w_{ij} = 0$ otherwise.

$$ATE_T^M = \frac{1}{N^T} \sum_{i \in T} [Y_i^T - \sum_{j \in C(i)} w_{ij} Y_j^C] \quad (6)$$

$$= \frac{1}{N^T} \sum_{i \in T} Y_i^T - \frac{1}{N^T} \sum_{j \in C} w_j Y_j^C \quad (7)$$

where the number of units in the treated group is denoted by N^T .

The variance of ATET

- Computing variance is not easy because the estimated propensity scores have variance themselves.
- There is no general indication among methods, some formulas work well in some cases, bootstrap in others:
- Abadie and Imbens (2008) showed that bootstrap is not a valid estimate for nearest-neighbor matching with continuous covariates.
- They propose formulas to compute variance for this estimator in a recent working paper (2011).
- More generally, matching estimators are intuitive but asymptotic properties are not always available.

Comments on matching methods

- The reliability of matching methods depends on the validity of the Unconfoundedness assumption. **BUT** This assumption is not directly testable.
- Existing tests are trying to assess indirectly its validity by testing the null hypothesis that an average causal effect is zero on sample where the particular average causal effect is known to equal zero (Rosenbaum (1987)):
- If **two control groups are available** (e.g. eligible and ineligible non-participants), compare their outcomes: this pseudo effect should be zero.
- If **pre-periods outcomes are available**, compare lagged outcomes for treatment and controls before policy: this pseudo effect should be zero.
- Samples used for controls matter: often problematic when treated and control samples are from very different datasets.

Implementing propensity score matching in R

For useful packages in R see:

stanford.edu/~ejdemyr/r-tutorials-archive/tutorial8.html

If the web-site does not work, search for standford.edu and then for R tutorial.