Matching Project

Benedetta Valpreda

March 2024

1 Introduction

In policy evaluation, we want to estimate the impact of a given policy on a certain outcome of interest.
To estimate such impact we would like to compare the outcome under the policy and what the outcome would have been without the intervention, but this is impossible to know. Therefore the primary challenge in policy evaluation is to find the appropriate counterfactual.

When there is not a clear assignment rule to the treatment (such as randomization or ranking), we turn to a method called Matching.
The underlying idea is that, for every treated unit, we search for the untreated unit (or a pool of them) that looks most similar based on the characteristics in the data set. Since the more available features the more difficult the matching (curse of dimensionality), instead of seeking untreated units with identical observable values, we aim to find untreated units with the closest probability of enrollment based on those features, which is known as the Propensity Score (PS).
In this way, for each treated unit, we can estimate the counterfactual with the closest untreated units in terms of PS. This allows us to compute the average difference in outcomes between treated and matched comparison units, providing an estimate of the intervention's impact.

One drawback of this method is that it is possible to match only according to observable features neglecting unobservable ones. If there exists an unobservable feature (which is correlated with the enrollment and is relevant for the outcome), omitting it from consideration can lead to biased estimates of the policy's impact.
If baseline data on the outcome are available, we can use such information to correct the bias. There are two possible way to do that:

1. perform Matching based on background characteristics (excluding baseline outcome levels) and use Difference-in-Difference (or DiD) on the treated and matched control groups. This means that, after PS-Matching, we compute the average difference in endline outcomes between the two groups,

the average difference in baseline outcomes and finally we calculate the difference between these two differences.

2. perform Matching based on both background characteristics and baseline outcome levels

In this project I show that the first method is able to correct the bias of any unobserved variable which is constant over time, while the second one reduces the bias but only in mean.

2 Case study

In this project, I imagine a policy evaluation application and I apply PS-Matching to estimate the policy impact.
Suppose we want to assess the impact of a training program on young individuals' wages in a given country. To take part in this program, people can enroll voluntary and we suppose that those with higher levels of motivation are more likely to apply. Therefore there exists a selection bias based on an unobservable variable (motivation). We suppose that motivation is constant over time.
The primary aim of this project is to estimate the impact of a such policy using the PS-Matching method, running Monte-Carlo simulations.

3 Data generation and Methodology

For the case study of this project, I generate a dataset of 1000 units and estimate the impact of the training program on wages. I compute the estimation with three different approaches:

1. **Method 1**: perform PS-Matching solely on background variables
2. **Method 2**: perform PS-Matching on background variables and combine it with DiD
3. **Method 3**: perform PS-Matching on both background variables and baseline outcome levels

I repeat this procedure multiple times (200) using different datasets to obtain impact estimations for each method across all datasets. Finally, I compute the mean estimated effects for each method to compare final results and draw conclusions.

For this analysis, the list of variables includes:

- **female**: binary variable that equals 1 for females and 0 for males
- **age**: numeric variable ranging from 18 to 25 years

- **latent education**: numeric variable from 0 to 1 indicating education score, with higher scores denoting higher education levels. I categorize this variable into three levels (1 to 3) representing low, medium and high educational levels.
- **motivation**: unobservable numeric variable that follows a normal distribution. This variable is used to generate outcome variables but is omitted when applying the three methods to estimate the impact of the program
- **treated**: binary variable that takes the value 1 for individuals participating in the program (those with motivation levels above the mean equal to 0) and 0 otherwise

The outcome variable of interest is wage. We first compute wage at baseline level as a function of individual explanatory variables (observable or not) plus a random error term ϵ (representing any "out of control" factor affecting wages like good/bad luck). As a result, in this project I generate baseline wage according to this equation:

$$wage_{pre} = 1500 - 100female + 20age + 10educationlatent + 20motivation + \epsilon \tag{1}$$

We suppose that the true impact of the training program is a wage increase of 200 units. Therefore I generate the wages at endline level with these equations:

$$\begin{cases} wage_{post} = wage_{pre} + 200 & \text{if } treated == 1 \\ wage_{post} = wage_{pre} & \text{otherwise} \end{cases}$$

Note that in practice the true treatment effect on the outcome is not known. Therefore we have to rely on the available data and estimate the impact from the information we have.

4 Results

This section shows the results obtained from running the MC simulations.
Across 200 simulations, I generate a new dataset, estimate the effect of the program using the three methods mentioned above and collect the resulting estimate for each method in every dataset. At the end of this procedure, for each method, I compute the mean estimate across all the simulations to obtain the final results, which are shown in the following table.

	mean impact
Method1	231.38
Method2	200
Method3	199.63

Table 1: Mean estimated impact across 200 MC simulations

We can see that Method1 (raw PS-Matching on background characteristics) provides a biased estimation of the treatment effect. Method2 (PS-Matching combined with DiD) is able to correct the bias and gives the exact true impact. Method3 (PS-Matching on background plus baseline outcome) closely approximates the true effect but only in mean.

5 Conclusions

This project is a valuable exercise to practice with the PS-Method.
Whenever we need to evaluate a policy without any specific rule of assignment we can match treated units with the most similar untreated units according to their probability of enrollment given the observable characteristics.
If there exists selection bias between treated and matched untreated units, the raw impact estimation derived from PS-Matching will be biased.
As long as we have data on outcome variable at baseline level, we can combine PS-Matching with DiD to take care of unobservable characteristics constant over time. This methods not only corrects the bias but also outperforms matching on both background features and baseline outcome, since this latter method corrects the bias only in mean (across several repetitions of the experiment).
In summary, the combination of PS-Matching and DiD offers a powerful approach to address selection bias and account for unobservable invariant factors in policy evaluation. By providing reliable estimates of policy impacts, this method can contribute to more informed decision-making processes in various policy settings.