Effect of Different Matching Methods on the Paper's Results

Julien Berger julien.berger@epfl.ch
Nicolas Bichon nicolas.bichon@epfl.ch
Johannes Brune johannes.brune@epfl.ch

Abstract

This report is based on the paper "Housing, Health, and Happiness" and aims to verify the robustness of the paper results. The balance of the treatment and control groups used in the paper is assessed and improved using propensity score and caliper matching methods. It was shown that balance improvements do not change the statistical significance of the paper results, which confirms paper authenticity.

1 Introduction

"Housing, Health, and Happiness" demonstrates that simple housing improvements in slum areas can be beneficial for child health as well as maternal satisfaction and happiness. If this is indeed true, the paper should be a source of inspiration for governments taking action to combat poverty. However, this pleasing conclusion is drawn from an observational study and may thus be biased (Stuart, 2010). For this reason, we verified if the same conclusion can be drawn once the paper treatment and control groups are well balanced. More precisely, we re-evaluated the treatment effect on 10 outcome variables. 5 variables related to cement floor coverage and 5 others to maternal satisfaction and happiness (tables 4 and 6 of the paper). The balance was done on 28 covariates using propensity score and caliper matching methods. We selected 5 types of variables, including housing, demographic, economic, hygienic and social covariates.

2 Data collection and description

The data used for the analysis are the same as those used in the paper. Data come from the file 'PisoFirme_AEJPol_20070024_household.dta' containing information at the household level. It includes data from both the 2000 Mexican Census and the 2005 Survey and contains 2'755 data points with complete information (1362 for the treatment and 1393 for the control).

3 Methods

In this section, we first define the propensity score and explain how we estimated its value. We then describe the matching algorithms used to balance treatment and control households and finally explain the methods used to visualize and assess this balance.

3.1 Propensity score

The propensity score is the probability of a unit receiving treatment given a set of observed covariates and is mathematically defined as

$$e(x) = Pr(T_i = 1|X_i)$$

where i indexes the unit, T is the treatment variable ($T = 1$ for treatment and $T = 0$ for control) and X is the vector of observed covariates. Using the propensity score to match samples is often simpler than using the whole set of covariates, as matching is done according to one scalar number instead of having to compare a large set of variables.

For each household included in the survey, propensity scores were predicted using logistic regression. This helped us modeling the binary treatment variable "dpisofirme" as a function of the 28 covariates and an intercept. Mathematically, the model can be written as

$$log(\frac{p}{1-p}) = b_0 + b_1 * x_1 + b_2 * x_2 + \ldots + b_28 * x_28$$

where p is the probability of the treatment variable being equal to one, b_0 represents the intercept and b_j is the coefficient of the j^{th} variable, $j = 1, ..., 28$. The propensity score could then be calculated, rewriting the above equation as

$$p = \frac{e^{b_0 * b_1 * ... * b_{28}}}{1 + e^{b_0 * b_1 * ... * b_{28}}}$$

3.2 Matching algorithm

Propensity scores were then used to find the similarity between each possible pair of treatment and control households, calculated as

$$similarity = 1 - |e(x)_{T=1} - e(x)_{T=0}|$$

With this relation, pairs with close propensity scores have a higher similarity with 1 being the maximal value. A bipartite graph was then built relating each possible pair of treated and control household with an edge weighted with the similarity calculated above. To find the optimal matching we used the maximum-weighted matching algorithm in which a matching is a subset of edges where no node occurs more than once. The weight of a matching is the sum of weights of its edges. Thus, we could find the best matching for which the propensity scores of each pair have highest similarity.

We finally tested the caliper matching method following the same algorithm, but taking into consideration a threshold (ϵ) of propensity score similarity. It includes only pairs that have a difference in propensity score less than ϵ to create the weighting of the bipartite graph. We tested different values ranging from 0.0001 to 0.05, resulting in smaller treatment and control sizes than for the first method not imposing any similarity threshold.

3.3 Balance visualization

Once a matching is made, it is necessary to assess the balance of the newly created groups. We used visualization tools to asses covariate balance. Quantitative discrete and continuous variables were visualized with histograms and box-plots. Histograms were supplemented by the corresponding kernel density functions. Binary covariates, on the other hand, were visualized with bar-plots showing the proportion of households having a value of 1. Bar-plots contain error bars representing bootstrapped 95% confidence intervals.

4 Results

The original matching corresponds to the 1362 treatment and 1393 control data points used in the paper. Visualization tools showed that the 28 covariates were relatively well balanced. However, some variables could be further balanced, notably S_headeduc (head of household's years of schooling), S_dem1 (proportion of males 0-5 years in household), S_hasanimals (binary variable indicating if a household has animals on land) and S_milkprogram (dummy variable indicating if a household is beneficiary of government milk supplement program). Results obtained by replicating tables 4 and 6 were all statistically significant at the 1% level.

4.1 Propensity score matching

The propensity score matching created two groups of same size (1362 treatment and control data points). Visualization tools showed that the covariates balance was extremely similar to the original matching. This is nevertheless not so surprising, since the number of data points in the new treatment group is the same as the original one (1362) and there is only a 31 difference in household number between the new control group and the original one (1362 instead of 1393). The treatment effect on the 10 outcomes of interest were re-evaluated (tables 4 and 6 of the paper) and the results were very similar to the paper. Indeed, all the results remained statistically significant at the 1% level.

4.2 Caliper matching with ϵ = 0.05

Caliper matching with $\epsilon = 0.05$ created two groups of same size (1349 treatment and control households). Visualization tools showed that the covariate balance was still extremely similar to the original matching. As a consequence, results of tables 4 and 6 were also very similar to the original tables and were still statistically significant at the 1% level.

4.3 Caliper matching with ϵ = 0.01

Matching with a stricter threshold ($\epsilon = 0.01$) created groups of vaguely smaller size (1191 treatment and control data points). This time, visualizing the 28 covariates revealed that the balance slightly improved for all of them, notably for S_headeduc, S_hasanimals, and S_milkprogram. This can clearly be seen on figure 1 showing the improvement in balance for the variable S_headeduc. In fact, we can see that the histograms as well as the box-plots of treated and

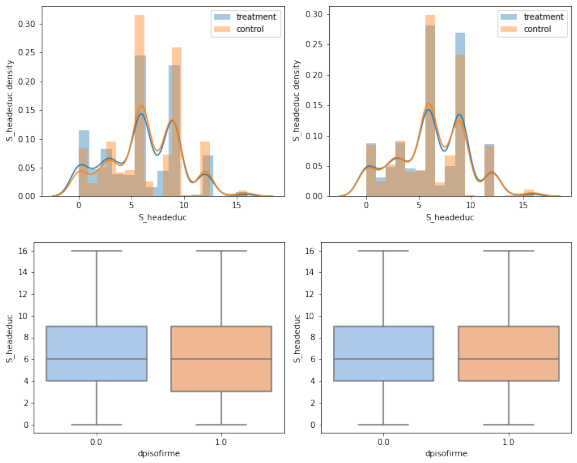

Figure 1: Histograms and box-plots for the variable S_headeduc before and after caliper matching. (left: original, right: caliper $\epsilon = 0.01$)

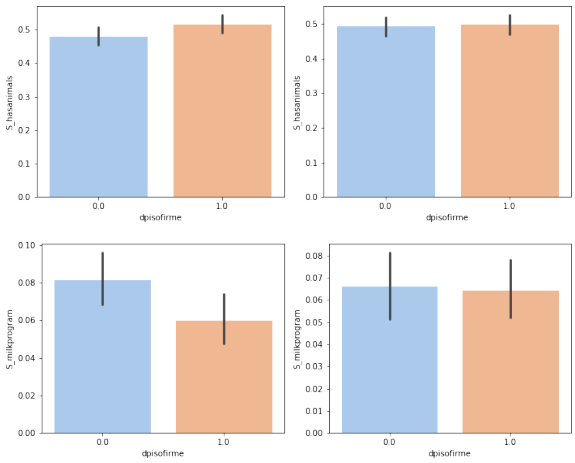

Figure 2: Bar-plots for the variables S_hasanimals and S_milkprogram before and after caliper matching. (left: original, right: caliper $\epsilon = 0.01$)

control groups show more similitude after applying the caliper matching method. Furthermore, improvements can also be seen on the bar-plots for the variables S_hasanimals and S_milkprogram represented on figure 2. Despite this overall balance improvement, re-evaluation of tables 4 and 6 gave results that were still all statistically significant at the 1% level.

4.4 Caliper matching with ϵ <0.01

Finally, we tried caliper matching with $\epsilon = 0.001$ and $\epsilon = 0.0001$ and obtained group sizes of 1103 and 586 households respectively. In both cases, the covariate balance was extremely similar to what was obtained with $\epsilon = 0.01$. This implies that reducing ϵ does not improve balance considerably. This is certainly due to the fact that the balance obtained with $\epsilon = 0.01$ is already close to perfect. As for $\epsilon = 0.01$, re-evaluation of tables 4 and 6 gave results that were all statistically significant at the 1% level.

5 Conclusions

In conclusion, "Housing, Health, and Happiness" is a paper which states that simple housing improvements can have beneficial impacts on child health and maternal happiness. Since this statement is drawn from an observational study, we decided to evaluate the treatment and control group balance to confirm the robustness of it. We demonstrated that these two groups were well balanced, which supports the paper results. We also improved the group balance using a caliper matching method with ϵ equal to 0.01 or less. All the Paper results remained statistically significant at 1% level, which supports even more the paper robustness. All in all, we think that this paper is extremely promising and politicians should refer to it when taking action against poverty.

References

Elizabeth A. Stuart. 2010. *Matching Methods for Causal Inference: A Review and a Look Forward.*