The Effect of Having a Daughter

Adrian M. Goedeckemeyer

Minerva Schools at KGI

The Effect of Having a Daughter

To investigate whether having a daughter has an effect on decision making of members in the US House of Representative we perform Genetic Matching on a Dataset acquired by sociologists in 2008. We are controlling for several discrete variables (race, gender, political party, and religion) and continuous variables (seniority, age and Democratic vote share). Total number of children, even though King included it, is not an appropriate variable to use, as it is directly related to our treatment variable.

We will not be including demographic data of the legislator's states (percentage females, college graduates), even though they are available in the dataset as they would introduce too many cofounders that are basically impossible to balance. A quick test showed incredible low p-values when including these. The democratic vote share of the state allows us to balance on some of the state's demographics summed up into one variable. We will perform sensitivity analysis on our best result to check how strong the influence of further covariates could be on our results.

The outcome variable measures the agreement of the representatives votes with positions of the National Organization of Women on a range from 0 to 100.

After a first run of Genetic Matching I achieved a mediocre balance, the lowest p-value was 0.15. In order to improve on that I added the square of age and of seniority as well as the interaction of democratic party affiliation and democratic voting preference in the state. These increased my p-value to about 0.33. To boost the p-value we had to add a stricter criterion on what could be matched using a caliper. The caliper defines the maximum matched distance for each variable in normed values. Especially variables that had low balance before received low calipers. Thus the total number of observations included in control is reduced, but the Balance on

our variables greatly improved. With about p=.6 Balance on the least balanced variable[1], I am satisfied with the results, it is a huge improvement from 0.002 on the unmatched data[2][3]. (Disclaimer: not all values are always exactly the same, GenMatch is not deterministic)

With this Balance achieved, I can evaluate the effect of having "anygirls" on voting and it appears that in my analysis there is a slight negative effect of -1.5766 with a SE of 0.77. The p-value for this effect is 0.042385, thus giving us high certainty on this effect[4]. To check the unconfoundedeness of these results I also perform a Rosenbaum Sensitivity test which gives us a low significance estimate of .3 of these results being due to unobserved factors. An increased Gamma of only 1.35 leads to only .75 alpha[5], thus greatly weakening our results, as such variables are very possible. Part of the reason for this high estimate is for one thing the low significance we start off with and the reduced dataset we received through the use of a caliper, which was necessary for balance, but also the nature of the problem as number of units in the 105th congress is low and such analysis requires Balance on many covariates to be meaningful, which reduces the amount of usable data.

We conclude that there is only a very minor measurable effect of having any girls on the observed voting behavior which should probably be neglected. The magnitude of the effect would definitely be miniscule compared to other factors such as party. I conclude that there is most likely an unobserved (in this analysis) variable that is responsible for much of the minor treatment effect we are seeing, as the Rosenbaum test declared as very likely.[6]

---

[1] See Table 2
[2] See Plot 1-6 in Appendix
[3] #significance: interprets significance of results, present throughout paper and especially the analysis in R
[4] See Table 1
[5] See Table 3
[6] #observation: This paper uses data from an observational data to extract accurate information on treatment effect and evaluates the results

Code References

Full Code with Output:

https://adriangoe.github.io/CS112/Washington_2008 [7][8]

Full code with Annotations but without output:

https://github.com/adriangoe/CS112/blob/master/Washington_2008.Rmd

Full code without annotations and no output:

https://github.com/adriangoe/CS112/blob/master/Washington_2008.R

Data References

https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/16598&widget=dataverse@king Gary King

[7] #simulation: Treatment effect is estimated through simulation of different models on our data. Attempt to test model dependency below was not very successful though. But an evaluation of the results is present in the paper in R Markdown document

[8] #algorithms: GenMatch uses a greedy algorithm to optimize weights assigned to different variables when matching

Graphs

Plot 1-6: https://adriangoe.github.io/CS112/Washington_2008#some-plots-to-illustrate-the-less-balanced-variables-before-and-after-matching

Plot 7-11: https://adriangoe.github.io/CS112/Washington_2008#looking-at-the-data

(for additional insights, beyond the main task of this paper)[9]

Tables

Table 1: https://adriangoe.github.io/CS112/Washington_2008#matching-results-3

Matching Results with Estimated Treatment Effect

Table 2: https://adriangoe.github.io/CS112/Washington_2008#balance-3

Balance Analysis for each Variable

Table 3: https://adriangoe.github.io/CS112/Washington_2008#sensitivity-analysis-for-the-last-results

Rosenbaum Sensitivity Test

[9] #dataviz: All plots are useful and worthy illustrations of the data