Simulation Exercise

Alexander Guggenberger

Course: Causal Inference

Summer Semester 2019

University of Vienna

1 Setting

I imagine somebody wants to investigate whether regularly drinking tea fosters longevity empirically. Thus, the setting I consider is the following:

D*	...	frequency of drinking tea
D	...	regularly drinking tea (binary)
distance	...	distance to English border
Z	...	being English
Y	...	life expectancy
V	...	some individual effects
X	...	measurable characteristics correlated with V

The assingment process is then described by the following equations:

$$distance \sim N(0,1) \tag{1}$$

$$Z = I_{distance \geq 0} \tag{2}$$

$$D^* = \beta * Z + \gamma * V \tag{3}$$

$$D = I_{D^* \geq 0} \tag{4}$$

For convenience, I interpret the assignment process as depending on personal characteristics V and, stereotypically, being English. I will later use distance to English border as an instrumental variable for being treated and as the running variable in an RDD estimation. By default, I assume positive correlation between V and U1 (0.7) and between V and U0 (0.1) The idiosyncratic gain U1 is higher for those who are treated (higher probability if V is high) and vice versa. This could be because those who benefit more from drinking tea are more likely to do it, eg. because they notice positive effects immediately. Also, there is positive correlation between U0 and V: even without drinking tea, tea drinkers would live longer, as they have a healthier lifestyle.

The true causal effect is +3 years.

Part 1		(1) ATE	(2) ATT	(3) ATNT	(4) LATE	(5) naive	(6) olsx
default	mean	3.016019	3.824332	1.95409	2.841493	4.156702	2.676225
	se	0.0086071	0.0116702	0.0116819	0.0330999	0.0111748	0.0171096
random	mean	3.016019	3.008503	3.026002	3.023061	3.024513	3.012681
	se	0.0086071	0.0127889	0.0116851	0.0353556	0.0127968	0.019989
distance important	mean	3.016019	3.329043	2.070219	2.068967	3.54614	2.200995
	se	0.0086071	0.0101132	0.0141923	0.017957	0.0136082	0.0139616
characteristics important	mean	3.014671	3.936389	2.05803	2.901773	4.274292	2.952847
	se	0.0085871	0.0118534	0.0108855	0.1183966	0.0114375	0.0178835

Part 2		(7) heck	(8) sharprdd1	(9) sharprdd2	(10) iv	(11) fuzzyrd	(12) psatt	(13) psate
default	mean	2.929979	0.3570281	0.4335794	2.9021	2.9021	2.876509	2.868892
	se	0.1006922	0.0209945	0.0470127	0.1135875	0.1135875	0.073896	0.0541048
random	mean	2.887076	0.3518917	0.4185687	2.979371	2.979371	3.048166	3.05908
	se	0.0961319	0.024009	0.0467302	0.1240859	0.1240859	0.0673674	0.0612165
distance important	mean	3.026418	1.00714	1.045719	2.093907	2.093907	2.558793	2.358895
	se	0.0290928	0.0202972	0.0433007	0.0252494	0.0252494	0.0391718	0.0282729
characteristics important	mean	8.714178	0.0226603	0.114357	4.262205	4.261989	2.956666	2.95134
	se	4.056298	0.0226059	0.0505925	2.546313	2.546238	0.1158313	0.0961689

2 Discussion

In the following I will firstly discuss the table of results row by row, and after that highlight which were the most noticeable insights to me.

The first row shows the result for the default setting as described above - the one also currently specified in the attached do-file. We see in the first three entries that the ATE is almost equal to the causal effect, which is not surprising given that idiosyncratic effects are 0 on average. However,

the above described selection leads to an ATT of 3.82 and an ATNT of 1.95, as the treated individuals have positive idiosyncratic effects (U1) on average, whereas the non-treated have negative ones in expectation. The LATE, the treatment effect on compliers is between ATE and ATNT, which makes sense, as those with really high idiosyncratic effects are always-takers and are therefore not included into the LATE.

As expected, the naive OLS estimator overestimates the effect: besides the causal effect, it does not only include the idiosyncratic positive U1 (which would lead to the ATT), but also the negative U0 of the untreated (as E(U0)=0 and V is positively correlated with U0, E(U0|D=0)<0). Controlling for X, however, even understates the effect. Under CIA, controlling for X is meant to lead to the ATE (Causal Effect in this case), but in this case, some of the positive causal effect seems to be wrongly attributed to X instead of the higher likelihood of being treated induced by X, leading to a smaller estimated effect of being treated. Being not only predictive for the assignment, but also for the output, X is not an optimal control. The Heckman estimator in column 7 (column 1 of the second part of the table), however, does the job pretty well, indicating the causal effect in the sense of ATE.

The next two columns report two attempted applications of sharp RD. Here, I assumed (wrongly) that being English deterministically explains the treatment status, which would allow me to interpret the random variable that we used to generate the binary assignment variable as the running variable. Sharprdd1 is the parametric RDD, where I fit a linear trend to below and above threshold outcomes (actually, there is no direct effect of D in this setting, so I would not even need to do this). Sharprdd2 is the non-parametric RDD, where I doubt that I can model the function below and above threshold correctly, so I instead restrict the sample to observations that are in the immediate neighbourhood of the threshold which allows me to ignore any effect of D on Y other than through D*. As the linearity assumption was not incorrect in my setting (no direct effect of D on Y $\implies$ zero slope, but linearity assumption still satisfied), there is no big difference in the coefficient, only the standard error gets larger, as I throw away data when I restrict the sample. However, both estimates are much too small, because I can in fact

not explain the treatment sufficiently by the assignment variable (X plays a big role). This highlights the importance of making the difference between a sharp and a fuzzy RD in case of doubt.

The next two columns shows the estimate of the fuzzy RD which is identical to an IV that uses the running variable as an instrument for being treated. In fact, I also used X for the first stage regression, as it is also exogenous, to improve the prediction of D. This estimator is a measure for the LATE, the effect on the compliers. It makes sense that it is a little smaller than the ATE, as those with the highest idiosyncratic treatment effects would have self selected into treatment anyway, therefore I am here measuring the effect on those with on average negative idiosyncratic effects. Actually, I would have expected standard errors to be larger for the iv, where I used the built in Stata function to do 2SLS, making Stata taking into account that predictions from the first stage are random variables, whereas they are taken as given if I run the first stage regression by hand like I did for column 11. However, they are identical.

The last two rows are the estimates for ATT and ATE using nearest neighbour matching. Note that the ATT is only slightly higher than the ATE. This is probably due to the fact that I had to take out extreme observations with respect to the propensity score st. matches could be found. As these where the observations with the most extreme Vs as well, I automatically took out all of the observations with extremely high positive idiosyncratic effects, leading to an underestimation of the causal effect (in the sense of the ATE as well as ATT). I do also not know the exact procedure that STATA uses to calculate the ATT, so maybe there is something wrong with that.

The second row shows the effect in a parallel universe where drinking tea is randomly assigned (and everyone complies), which I simulated by eliminating any correlation between V and U1 and U0. Clearly, ATE, ATT, ATNT and LATE coincide in such a setting, and all the estimation methods get very close to the true value for the causal effect, except the two sharp RDD estimations. On one hand, I do not have to deal with any correlation and selection, but still the assignment to treatment depends on V quite strongly, and assuming an RDD design fails to capture that.

In order to obtain the values in the third row, I changed the assignment process by setting β, the coefficient for distance, to 10. This means, that distance becomes the determining factor for being treated, whereas V, the part correlated to the idiosyncratic effects, becomes less important relatively. Consequently, the difference between ATE, ATT and ATNT decreases and a naive OLS estimation is closer to the causal effect. The OLS with controls underestimates the causal effect even more than in the above rows, as the importance of X on D is supposed to be bigger than it is, so I am wrongly attributing even a higher fraction of the variation in Y to X instead of D. Sharp RDD estimation works better, as the assumption of a 100% jump in the likelihood of being treated is closer to the real setting. The LATE, measured by the IV regression (equivalently the fuzzy RDD) becomes smaller, because the predictive power of X for the treatment assignment decreased, whereas it still predicts the idiosyncratic effect, thus the actual outcome, as well as before. (The causality of D* is "bypassed".)

The two nearest neighbour matching estimators got even farther off. This is probably due to a higher variation in propensity score, so I omit even more of the observations with high treatment effects.

The fourth row reports the opposite case, I reset β=1 and instead γ=10, i.e. treatment assignment depends strongly on individual characteristics, i.e. those with high idiosyncratic effects are very likely to end up treated. This leads to a high ATT and naive OLS, sharp RDD of course fails. The OLS regression with controls does pretty well in abstracting from the selection problem and finding the ATE, as indeed V (and the correlated X) is mainly responsible for the assignment. The Heckmann estimator suggests a much too high causal effect, so using controls is the better option if very strong controls are available, like in this case. IV/fuzzy RDD reports a very high estimate, which I cannot fully explain, it should be the LATE (but the standard error is pretty high).

The propensity score ATE is close to the real one. The propensity score ATT, however, is far too low, it is hardly even distinct from the ATE. This might be because here, propensity scores depend highly on X, as X determines D so strongly. Thus, by ignoring high propensity scores (as they have no

match in the control group) I ignore high idiosyncratic effects, and I am especially ignoring always-takers, who have high score and high X.

3 Conclusion

In conclusion, the most important findings are the following:

- A naive OLS estimation does neither estimate the ATE nor the ATT if there is self selection, as it includes both U1 and U0 of the treated, ie. also counter-factual idiosyncratic effects matter.
- Propensity score based nearest neighbour matching fails to calculate ATT if only a very limited range of the propensity score has support in both groups, because this means that the typical treated individual is excluded from the estimation procedure.
- The example of the sharp RDD highlights the importance of checking whether an approach really fits the setting - if the assignment the treatment does not depend on the running variable deterministically, but only stochastically, sharp RDD cannot be used.
- An IV/fuzzy RDD estimation identifies the LATE reliably only if the instrument does not have a strong direct effect on (or correlation with) the outcome. That's why it worked well when distance was important for the treatment assignment (i.e. being the dominant instrument), whereas it failed to identify the LATE efficiently when V was the important factor (meaning that X, now the dominant instrument, was also predicting the output well, not only the assignment to treatment).
- The same is true for OLS with controls. As X also explains the outcome well (positive correlation of X through V with assignment but also with U1 and U0) including it into the regression leads to a downwards bias as some of the treatment effect is attributed to X wrongly.