MASTERS IN COMPUTATIONAL SOCIAL SCIENCE
THE UNIVERSITY OF CHICAGO

Extrapolation of Treatment Effect in Sharp Regression Discontinuity Design

Nan Ge
The University of Chicago

ABSTRACT

Canonical RD estimation identifies a local average treatment effect, but has limited external validity.

This paper proposes machine learning methods to extrapolate the average treatment effect based on a regression discontinuity design.

With non-unrealistic assumptions, double LASSO performs very well in extrapolation. However, more progress is needed in asymptotic inferences.

INTRODUCTION

- Regression discontinuity (RD) design is one of the most credible identification strategy in the absence of a randomiezd experiment (Cattaneo et al. 2018).
- Hahn, Todd and Van der Klaauw (2001) shows that under the assumption that E[Y(d)|R=c],where d ∈ {0,1}, are continuous at the treatment threshold c, E[Y(1)-Y(0)|R=c] is identifiable.
- Therefore, canonical RD parameter identifies a local average treatment effect, therefore, has limited external validity.
- In this paper, a double machine learning method is provided for the extrapolation of RD effects in sharp RD design.
- Adapt the double machine learning method in Chernozhukov et al. (2018) to extrapolate the RD treatment effect.
- Prove the underlying behavioral assumption in CIA in Angrist and Rokkanen (2015).
- Use double machine learning approach to select relevant covariates that make CIA works. The method can be extended to multiple-cutoffs RD designs.
- The simulation results show that the double machine learning approach performs better than simple approach in Angrist and Rokkanen (2015).
- Contributions to the existing literature:
  - Angrist and Rokkanen (2015): more principled way of variable selection.
  - Cattaneo et al. (2019): extend their identification results to the case when there are more than two cutoffs in RD design.

IDENTIFICATION METHOD

Set up

- Running variable: R_i; binary treatment: D_i;; and potential outcomes: Y(d), where d ∈ {0,1}.
- Treatment assignment rule: D = 1{R>c}
- A rich set of covariates: X is high-dimensional

Assumptions

- Conditional Independence Assumption (CIA): E[Y(d)|R, X] = E[Y(d)|X]
- Common Support Assumption (CSA): 0 < P[D_i = 1 | X_i] < 1
- Angrist and Rokkanen (2015) proved that ATE is identifiable under these assumptions.

Process

1. Split the sample into two folds: train set and test set.
2. Use Angrist and Rokkanen's method/LASSO/double LASSO select a set of covariates Z
3. Use the test set to estimate E[Y(d)|Z] by estimating E[Y|Z, D=d]
4. Use test set to compute the extrapolated ATE by E[Y(1) – Y(0)] = E[E[Y(1) – Y(0)|Z]]
5. Bootstrap the standard error of extrapolated ATE

DATA GENERATING PROCESS

- Simple cutoff
  - Case 1: Parameters of similar scale
    $y_i(0) = 0.8 + 18.3x_{1i} - 13.3x_{2i} - 15.6x_{3i} + \varepsilon_{0i}$
    $y_i(1) = 1.5 + 23.5x_{1i} - 21.6x_{2i} + 22.6x_{3i} + \varepsilon_{1i}$
    $r_i = 1.6 + 6.8x_{1i} + 8.3\, x_{2i} + 7.6\, x_{3i} + u_i$
  - Case 2: Parameters of different scale
    $y_i(0) = 0.8 + 2.3x_{1i} - 6.3x_{2i} - 15.6x_{3i} + \varepsilon_{0i}$
    $y_i(1) = 1.5 + 23.5x_{1i} + 1.6x_{2i} + 2.6x_{3i} + \varepsilon_{1i}$
    $r_i = 1.6 + 6.8x_{1i} + 3.3\, x_{2i} + 3.6\, x_{3i} + u_i$
- Multiple cutoff
  - Case 3: Same type
    Similar as case 1 and 2, but with cutoff 2 and 5
  - Case 4: Different types
    $y_{i0}(1) = 0.65 + 9.6x_{1i} - 3.6x_{2i} - 3.5x_{3i} + \varepsilon_{0i}$
    $y_{i1}(1) = 1.5 + 23.5x_{1i} + 1.6x_{2i} + 2.6x_{3i} + \varepsilon_{1i}$
    $y_{i0}(2) = 0.5 + 8.6x_{4i} - 5.5x_{5i} - 3.5x_{6i} + \varepsilon_{0i}$
    $y_{i1}(2) = 1.3 + 19.9x_{4i} + 3.6x_{5i} + 8.8x_{6i} + \varepsilon_{1i}$
    $r_{i1} = 1.6 + 5.6x_{1i} + 8.3x_{2i} + 6.6x_{3i} + u_i$
    $r_{i2} = 2.3 + 6.5x_{4i} + 8.5x_{5i} + 5.3x_{6i} + u_i$

RESULTS

Figure 1. Example of simulated dataset with single cutoff

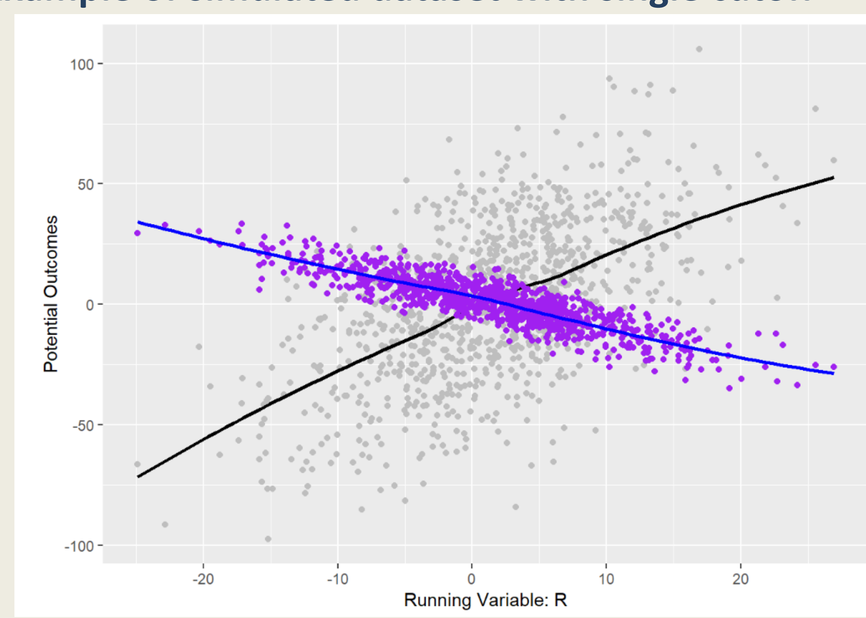

Table 1. Extrapolation results for case 1

Methodology	AR + Split	AR + LASSO	AR + DML
Extrapolation Error	0.5159	0.04276	-0.08795
Abs(Extrapolation Error)	1.905466	2.06358	1.44159
MSE of Extrapolation Error	10.30503	6.20552	3.141086
N	1000	1000	1000
Extrapolation Error	0.1624	0.164	0.1639
Abs(Extrapolation Error)	0.97363	0.972988	0.972924
MSE of Extrapolation Error	1.487618	1.483372	1.483115
N	2000	2000	2000

Table 2. Extrapolation results for case 2

Methodology	AR + Split	AR + LASSO	AR + DML
Extrapolation Error	0.1665	3.7291	0.01781
Abs(Extrapolation Error)	0.799501	3.7404	0.674
MSE of Extrapolation Error	1.072785	15.16516	0.693618
N	1000	1000	1000
Extrapolation Error	0.07641	3.8978	0.07721
Abs(Extrapolation Error)	0.559227	3.991	0.560756
MSE of Extrapolation Error	0.4676639	17.1152	0.4679712
N	2000	2000	2000

RESULTS (Cont.)

Table 3. Extrapolation results for case 3

Cutoff 1			
Methodology	AR + Split	AR + LASSO	AR + DML
Extrapolation Error	-0.004317	3.619	0.002138
Abs(Extrapolation Error)	0.506142	3.619	0.6932
MSE of Extrapolation Error	0.3580091	13.835	0.714849
N	2000	2000	2000
Cutoff 2			
Methodology	AR + Split	AR + LASSO	AR + DML
Extrapolation Error	0.01658	4.471	-0.1401
Abs(Extrapolation Error)	0.482994	4.471	0.81516
MSE of Extrapolation Error	0.3478324	20.966	0.986248
N	2000	2000	2000

Table 4. Extrapolation results for case 4

Cutoff 1			
Methodology	AR + Split	AR + LASSO	AR + DML
Extrapolation Error	1.0139	4.293	0.04702
Abs(Extrapolation Error)	1.248002	4.293	0.332591
MSE of Extrapolation Error	7.9508	18.58	0.157128
N	2000	2000	2000
Cutoff 2			
Methodology	AR + Split	AR + LASSO	AR + DML
Extrapolation Error	1.3691	11.978	0.029917
Abs(Extrapolation Error)	1.717539	11.978	0.495368
MSE of Extrapolation Error	10.2483	144.42	0.3658364
N	2000	2000	2000

Table 5. Bootstrapped coverage rate

DGP 1			
Methodology	AR + Split	AR + LASSO	AR + DML
Coverage Probability	1	1	1
DGP 2			
Methodology	AR + Split	AR + LASSO	AR + DML
Coverage Probability	1	0.99	1

DISCUSSION & CONCLUSIONS

- Double LASSO performs the best, especially in multiple cutoff cases, and when the parameters differ in scale. Naïve LASSO performs the worst.
- With machine learning techniques, we could extrapolate ATE, thus improving the external validity of regression discontinuity designs.
- However, bootstrap is not suitable for asymptotic inferences. Our bootstrap experiment suggests that the 95% confidence interval of extrapolation covers the true ATE with 100% probability.

REFERENCES

Angrist, Joshua D., and Miikka Rokkanen. "Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff." *Journal of the American Statistical Association* 110.512 (2015): 1331-1344.

Chernozhukov, Victor, et al. "Double/debiased machine learning for treatment and structural parameters." (2018): C1-C68.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. "A Practical Introduction to Regression Discontinuity Designs: Volume I." (2018).

CONTACT

Nan Ge
Master of Computational Social Science, The University of Chicago
Email: genan@uchicago.edu
Website: https://github.com/jsgenan