Economics 675: Applied Microeconometrics – Fall 2018

Assignment 6 – Due date: Mon 10-Dec

Last updated: June 4, 2018

Contents

Guidelines:

- You may work in (small) groups while solving this assignment.
- Submit individual solutions via `http://canvas.umich.edu` in one PDF file collecting everything (e.g., derivations, figures, tables, computer code).
- Start each question on a separate page. Always add a reference section if you cite other sources.
- Clearly label all tables and figures, and always include a brief footnote with useful information.
- Always attach your computer code as an appendix, with annotations/comments as appropriate.
- Please provide as much detail as possible in your answers, both analytical and empirical.

1 Question 1: Continuity-Based Identification in SRD Designs

This question discusses formal (nonparametric) continuity-based identification results for sharp RD (SRD) designs, following the results in Hahn, Todd, and van der Klaauw (2001), Lee (2008) and Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016).

Consider the following multi-cutoff RD framework:

- X_i denotes the running variable (or score) for unit i, with continuous density $f_X(x)$.
- C_i denotes the cutoff faced by unit i, with $\mathbb{P}[C_i = c] \in [0,1]$ for $c \in \mathcal{C} = \{c_1, c_2, ..., c_J\}$.
- $T_i = T_i(X_i, C_i)$ denotes the observed treatment status for unit i.
- $Y_{1i}(c)$ and $Y_{0i}(c)$ denote, respectively, the potential outcomes under treatment and control for each cutoff level $c \in \mathcal{C}$. Therefore, the observed outcome is

$$Y_i = Y_{1i}(C_i) \cdot T_i + Y_{0i}(C_i) \cdot (1 - T_i).$$

Throughout this exercise we assume regularity conditions hold, such as densities existing, being bounded and bounded away from zero, or other assumptions required to interchange limits and integral operators.

In SRD Designs, treatment compliance is perfect or, alternatively, intention-to-treat parameters are of interest. As discussed in class, a common empirical practice is to define a normalized score $\widetilde{X}_i := X_i - C_i$, which pools all the observations as if there was only one RD cutoff at $\widetilde{X}_i = 0$. In this setting, $T_i = T_i(X_i, C_i) = \mathbf{1}(X_i \geq C_i) = \mathbf{1}(\widetilde{X}_i \geq 0)$ for all units in the sample.

We impose the following assumptions, for all $c \in \mathcal{C}$:

(S1) $\mathbb{E}[Y_{0i}(c)|X_i = x, C_i = c]$ and $\mathbb{E}[Y_{1i}(c)|X_i = x, C_i = c]$ are continuous in x at $x = c$.
(S2) The conditional density of $X_i|C_i$, denoted $f_{X|C}(x|c)$, is continuous in x at $x = c$.
(S3) $\lim_{\varepsilon \to 0^+} \mathbb{E}[T_i|X_i = c + \varepsilon, C_i = c] = 1$ and $\lim_{\varepsilon \to 0^+} \mathbb{E}[T_i|X_i = c - \varepsilon, C_i = c] = 0$.

Using these assumptions, answer the following questions.

1. Assume $\mathbb{P}[C_i = c] = 1$. Show that

$$\tau_{\texttt{SRD}} = \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i|\widetilde{X}_i = \varepsilon] - \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i|\widetilde{X}_i = -\varepsilon] = \mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i = c]$$

   Compare this result to Hahn, Todd, and van der Klaauw (2001).

2. Assume $\mathbb{P}[C_i = c] \in (0,1)$ for all $c \in \mathcal{C}$. Show that

$$\begin{aligned}\tau_{\texttt{SRD}} &= \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i|\widetilde{X}_i = \varepsilon] - \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i|\widetilde{X}_i = -\varepsilon], \\ &= \sum_{c \in \mathcal{C}} \mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i = c, C_i = c] \cdot \frac{f_{X|C}(c|c)\mathbb{P}[C_i = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i = c]}.\end{aligned}$$

   Give an intuitive interpretation of this result, and explain how the single-cutoff and the multi-cutoff sharp RD designs are related. Under which conditions, both settings estimate the same population parameter?

3. Assume $\mathbb{P}[C_i = c] \in (0,1)$ for all $c \in \mathcal{C}$. Consider the following simple model of heterogeneity: $Y_{ti}(c) = y_t(c, W_i)$, where $y_t(\cdot)$ is a fixed map, $t = 0, 1$, $c \in \mathcal{C}$, and W_i are i.i.d. draws representing "unobserved individual characteristics". Show that

$$\begin{aligned}\tau_{\texttt{SRD}} &= \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i | \widetilde{X}_i = \varepsilon] - \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i | \widetilde{X}_i = -\varepsilon], \\ &= \sum_{c \in \mathcal{C}} \int (y_1(c, w) - y_0(c, w)) \cdot \frac{f_{X|C,W}(c|c, w)}{f_{X|C}(c|c)} \cdot \frac{f_{X|C}(c|c)}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c) \mathbb{P}[C = c]} \cdot \mathbb{P}[C = c | W_i = w] \cdot F_W(\mathrm{d}w).\end{aligned}$$

Compare this result to Lee (2008). Give an intuitive interpretation of this result, and explain how the single-cutoff and the multi-cutoff sharp RD designs are related. Under which conditions, both settings estimate the same population parameter?

2 Question 2: The Effect of Head Start on Child Mortality

This question investigates the effect of Head Start on Child Mortality using the dataset and original empirical RD approach proposed by Ludwig and Miller (2007). The main goal is to conduct a complete RD empirical analysis, including different graphical methods (RD plots), alternative falsification methods, and appropriate inference methods. The questions below follow closely the methodological and empirical results in Cattaneo, Titiunik, and Vazquez-Bare (2017), which compares different inference approaches for RD designs.

The dataset used is `HeadStart.dta`. See the Appendix (Section 3 below) for details.

2.1 RD Plots and Falsification Tests

1. Construct four RD plots (of the pre-intervention variable `mort_related_pre`), using both (i) evenly-spaced and (ii) quantile-spaced binning, and choosing the total number of bins either (a) to be IMSE-optimal or (b) to mimic the overall data variability.
   Compare the results and explain how this graphical devices can be used for falsification purposes.

2. Conduct formal falsification of the RD design, using (i) histogram plots, (ii) local binomial tests, and (iii) continuity-in-density tests.
   Compare the results and explain how this formal methods can be used for falsification purposes.

Additional falsification tests can be conducted using "placebo" analysis (on "pre-intervention" covariates and "post-intervention" outcomes not affected by treatment) and "robustness" methods ("donut hole", "placebo cutoff" and "bandwidth change" analysis), as discussed in class. You will be asked to conduct these additional falsification tests below.

2.2 Global and Flexible Parametric Methods

These methods are not recommended for practice, but it is useful to try them out at least once.

1. Assuming a "constant treatment effect model" (i.e., additive separable constant effect), estimate the RD treatment effect using a p-th order **global** polynomial with $p = 3, 4, 5, 6$.
   Plot the fitted values.
   Compare the results and explain why this method is unlikely to provide reliable results in practice.

   Table: Global polynomial fit under constant treatment effect assumption

p	3	4	5	6
Point estimate				
Standard error	( )	( )	( )	( )

2. Assuming an "heterogeneous treatment effect model" (i.e., fully interacted effect), estimate the RD treatment effect using a p-th order **global** polynomial with $p = 3, 4, 5, 6$.
   Plot the fitted values.
   Compare the results and explain why this method is unlikely to provide reliable results in practice.

   Table: Global polynomial fit separately on the two sides of the cutoff

p	3	4	5	6
Point estimate				
Standard error	( )	( )	( )	( )

3. Assuming a "local parametric model", estimate the RD treatment effect using a p-th order **local** polynomial with $p = 0, 1, 2$ for ad-hoc bandwidths choices $h = 1, 5, 9, 18$.
   Plot the fitted values.
   Compare the results and explain why this method is unlikely to provide reliable results in practice.

Table: Local parametric model

$h = 1$

p	0	1	2
Point estimate			
Standard error	( )	( )	( )

$h = 5$

p	0	1	2
Point estimate			
Standard error	( )	( )	( )

$h = 9$

p	0	1	2
Point estimate			
Standard error	( )	( )	( )

$h = 18$

p	0	1	2
Point estimate			
Standard error	( )	( )	( )

4. Are the results above consistent with each other? Explain under which conditions one would expect this to be the case.

2.3 Robust Local Polynomial Methods

Consider now fully nonparametric methods based on local polynomial techniques.

1. Construct MSE-optimal RD point estimators and robust confidence intervals using local constant, local linear and local quadratic estimators ($p = 0, 1, 2$).
   Compare the results and explain why this method is asymptotically valid, and optimal in same cases.

Table: local polynomial estimation

$p = 0$, local constant

	Point estimate	Standard error	Confidence interval
Conventional		( )	[,]
Bias-corrected		( )	[,]
Robust		( )	[,]

$p = 1$, local linear

	Point estimate	Standard error	Confidence interval
Conventional		( )	[,]
Bias-corrected		( )	[,]
Robust		( )	[,]

$p = 2$, local quadratic

	Point estimate	Standard error	Confidence interval
Conventional		( )	[,]
Bias-corrected		( )	[,]
Robust		( )	[,]

2. Conduct the following robustness checks on your findings.

a) Conduct hypothesis tests of zero treatment effect on placebo outcomes, using both "pre-intervention" covariates and "post-intervention" outcomes not affected by treatment.

b) Employ different bandwidth and kernel choices, and re-compute the main inference results for the RD treatment effect.

Table: robustness check (different bandwidths and kernel functions)

(local linear regression $p = 1$ and bias-corrected estimates)

Bandwidth $h =$	1	2	3	4	5	6	7	8	9	10
Triangular										
Uniform										
Epanechnikov										

c) Employ a "donut hole" approach, and re-compute the main inference results for the RD treatment effect excluding the closest $\ell = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ observations to the cutoff.

Table: robustness check (drop ℓ observations nearest to the cutoff)

(local linear regression $p = 1$ and bias-corrected estimates)

#(obs) dropped $\ell =$	1	2	3	4	5	6	7	8	9	10
Point estimate										

d) Employ a "placebo cutoff" approach, and re-compute the main inference results for the RD treatment effect using cutoffs $c = -10, -8, -6, -4, -2, 2, 4, 6, 8, 10$. Do you expect to find significant treatment effect?

Table: robustness check (different cutoff c)

(local linear regression $p = 1$ and bias-corrected estimates)

cutoff $c =$	-10	-8	-6	-4	-2	2	4	6	8	10
Point estimate										
p-value										

e) Extra credit: using the results in Calonico, Cattaneo, and Farrell (2018), re-compute the main inference results for the RD treatment effect using coverage error rate (CER) optimal bandwidth (as opposed to a MSE-optimal bandwidth). See Cattaneo and Vazquez-Bare (2016) for more discussion.

3. Based on the empirical findings above, give a complete assessment of the effect of Head Start on Child Mortality based on continuity-based RD methods.

2.4 Local Randomization Methods

Consider now local randomization methods based on analysis of experiments techniques.

1. Find a neighborhood around the cutoff where the local randomization assumption is plausible, using the method based on balance tests on "pre-intervention" covariates and "post-intervention" outcomes not affected by treatment. Employ different statistics to assess the sensitivity of this finding.
   Produce RD plots of the outcome variables for the selected neighborhood using evenly-spaced binning with IMSE-optimal number of bins.

2. Using the neighborhood selected above, employ local randomization (Neyman's and Fisher's) approaches to study the the effect of Head Start on Child Mortality.
   Give a detail explanation of the interpretation of point estimators, hypothesis test and confidence intervals in this setting. Make sure to explain the differences between asymptotic approximations and finite-sample exact results.

3. Explore the sensitivity of the results by changing the neighborhood around the cutoff where the local randomization assumption is assumed to hold.
   Extra credit: using Rosenbaum's sensitivity analysis approach, assess the robustness of the empirical findings above; explain how this method work, and give a precise interpretation of the findings.

Table: sensitivity analysis (changing the window size w)

(Neyman's approach)

Window size $w =$	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6
Point estimate										
Standard error	( )	( )	( )	( )	( )	( )	( )	( )	( )	( )
p-value										

4. Based on the empirical findings above, give a complete assessment of the effect of Head Start on Child Mortality based on Local-Randomization-based RD methods.

3 Appendix: headstart.dta Data Description

The focus here is to study the effect of Head Start assistance on child mortality by RD design. Head Start was established in 1965 as part of the War on Poverty to provide preschool, health, and other social services to poor children age three to five and their families. Interest in Head Start is motivated in part by large disparities in cognitive and noncognitive skills along race and class lines observed well before children start school, and by arguments that human capital interventions may be particularly promising for disadvantaged children during the early years of life.

Here we exploits a discontinuity in program funding across counties that resulted from OEO's implementation of the program. Specifically, in order to ensure that applications from the poorest communities would be represented in a nationwide grant competition for the program's funds, the federal government provided assistance to the 300 poorest counties in the U.S. to write and submit applications for Head Start funding. This led to increased Head Start participation and funding rates in these counties, creating a discontinuity in program participation at the 300th poorest county. For details see Ludwig and Miller (2007).

The dataset `HeadStart.dta` contains 2,783 observations and four variables:

(i) Outcome variable (`mort_related_post`): county-level mortality rates of children five to nine years of age over the period 1973–1983 due to causes addressed as part of Head Start's health services (measured as one-year mortality rates per 100,000 Head Start-related causes).

(ii) Running variable (`povrate60`): county-level poverty index in 1960, normalized so that the cutoff is 0 and a county is treated (received assistance) if and only if the poverty index is nonnegative.

(iii) Pre-intervention variable (`mort_related_pre`): similar to `mort_related_post`, but is measured before the implementation of Head Start.

(iv) `mort_injury_post`: county-level mortality rates of children five to nine years of age over the period 1973–1983 due to injury. Although measured after the implementation of Head Start, it is believed to be unaffected by the program.

Table 1: Summary Statistics

Variable	Mean	St.Dev.	Min	Median	Max
Running Variable					
`povrate60`	-22.464	15.268	-43.990	-25.606	22.372
Outcome					
`mort_related_post` (all)	2.254	5.726	0.000	0.000	136.054
`mort_related_post` (treat)	2.422	4.513	0.000	0.000	29.895
`mort_related_post` (control)	2.234	5.853	0.000	0.000	136.054
Other					
`mort_related_pre` (all)	7.219	17.197	0.000	3.109	576.441
`mort_related_pre` (treat)	10.517	14.226	0.000	7.312	91.575
`mort_related_pre` (control)	6.830	17.476	0.000	2.484	576.441
`mort_injury_post` (all)	21.182	17.177	0.000	18.761	215.054
`mort_injury_post` (treat)	25.960	19.655	0.000	22.334	168.705
`mort_injury_post` (control)	20.618	16.775	0.000	18.429	215.054

All: whole sample. **Treat**: the treatment group, defined by running variable being nonnegative (sample size 294). **Control**: the control group, defined by running variable being negative (sample size 2489).

References

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018): "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, forthcoming.

Cattaneo, M. D., L. Keele, R. Titiunik, and G. Vazquez-Bare (2016): "Interpreting Regression Discontinuity Designs with Multiple Cutoffs," *Journal of Politics*, 78(4), 1229–1248.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2017): "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality," *Journal of Policy Analysis and Management*, 36(3), 643–681.

Cattaneo, M. D., and G. Vazquez-Bare (2016): "The Choice of Neighborhood in Regression Discontinuity Designs," *Observational Studies*, 2, 134–146.

Hahn, J., P. Todd, and W. van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 201–209.

Lee, D. S. (2008): "Randomized Experiments from Non-random Selection in U.S. House Elections," *Journal of Econometrics*, 142(2), 675–697.

Ludwig, J., and D. L. Miller (2007): "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *Quarterly Journal of Economics*, 122(1), 159–208.