Problem Set 7

Quantitative Political Methodology (L32 PS 363)

Instructions

1. Print out and write your WUSTL ID at the top of **each** page, and complete each exercise in the space allotted. You may attach extra pages if the space provided is not sufficient, but please indicate that you have done so below.

   Pages attached:_____

2. Please **show your work** if possible. You may lose points by simply writing in the answer. If the problem requires you to use `R`, please include the code you used to get your answers. If you are not sure if work needs to be shown for a particular problem, please ask a TA or post a question on Facebook.

3. The various pages of your homework should be **stapled together** (no paper clips please). If pages are lost because of a lack of a staple, no credit will be granted for that portion of the homework.

4. This problem set is **due at the beginning of class on Wednesday, November 29, 2017**. No late assignments will be accepted.

5. Total available points for this homework: 40.

Question 1 (Total: 14 points)

Longo et al. (2014) investigate whether checkpoints (which are nonviolent impediments to movement, such as roadblocks and gates) in the West Bank make Palestinians less supportive of violence against the Israeli population.[1] To answer their research question, the authors exploit a policy intervention on May/June 2009 that eased the Za'atara checkpoint. Meanwhile, the Wadi Nar checkpoint did not undergo any change. The authors conducted a survey on people living in villages near the two checkpoints, before and after the intervention in 2009. With the data they collected, the authors do difference-in-difference estimation to find out the effect of checkpoint easement on Palestinian political attitudes.

(a) (2 points) Since survey data are at the individual level, difference-in-difference estimation can be done with a regression. The formula is: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 * X_2) + \epsilon$. Write out the equation in the context of this study. That is, rewrite the formula filling in what X_1 and X_2 are in Longo et al.'s study.

(b) (4 points) Explain what β_1, β_2, and β_3 would capture in the context of the study, according to your answer in part (a).

[1] Longo, Matthew, Daphna Canetti, and Nancy Hite-Rubin. 2014. "A Checkpoint Effect? Evidence from a Natural Experiment on Travel Restrictions in the West Bank." *American Journal of Political Science* 58(4): 1006-1023.

(c) (4 points) Download `https://jmontgomery.github.io/ProblemSets/longo.csv`. It is a simplified version of Longo et al.'s data. There are three variables:

`militancy` measures Palestinians' attitudes toward violence. Higher values mean more support for violence.

`Sample2009` is a binary variable that takes 1 if the survey was conducted after the intervention and 0 if before.

`ZA` is also a binary variable. It takes the value of 1 if the survey respondent is from a village near Za'atara, and 0 if the respondent is from a village near Wadi Nar.

Import the data into `R`. Run a regression to replicate Longo et al.'s difference-in-difference estimation. Report and interpret your DID estimate. Does checkpoint easement make Palestinians less militant?

(d) (4 points) Create a plot (either in `R` or by hand) that shows the difference-in-difference. Make sure you include a dotted line for the counterfactual and clearly label all parts of the plot.

Question 2 (Total: 14 points)

In political science, gender gap in political representation and participation is an important topic of study. Suppose that you have data on the 2002 state legislative elections. The dataset contains information on districts where a female candidate ran against a male candidate. For the districts that are included in the 2002 dataset, you also have female voter turnout in the subsequent 2004 election. In sum, suppose you have three types of information:

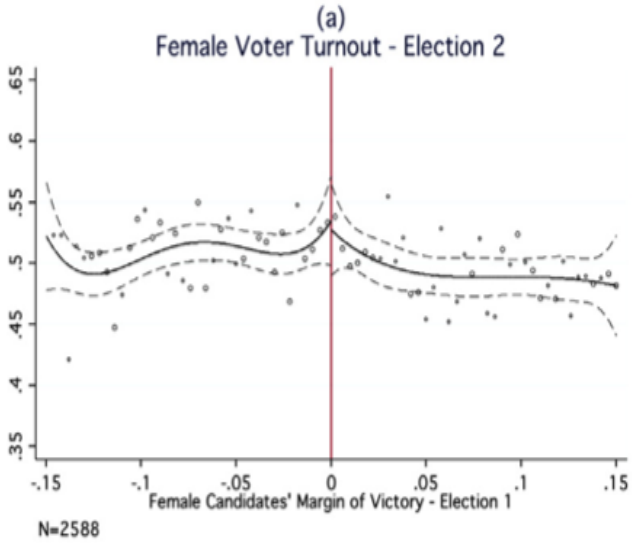
- The vote share a female candidate received in the district in the 2002 election.
- Whether the female candidate won the district in the 2002 election.
- Female voter turnout in the district in the 2004 election.

(a) (4 points) With the data you have, you can use regression discontinuity estimate the effect of electing a woman to office on female voter turnout in the next election. What is the outcome we are interested in? What is the treatment? What is the threshold of the treatment?

(b) (4 points) The equation for a regression discontinuity design is: $y = \beta_0 + \beta_1 X + \beta_2 * I(X > D) + \epsilon$. Identify what y, X, D, and $I(X > D)$ are in your study.

(c) (3 points) What is β_2 estimating?

(d) (3 points) Suppose that someone did (a more complicated) RD estimation for you to answer this question and gives you the following plot.[2] What is the plot telling us? That is, what is your conclusion about the effect of female representation on female turnout? Explain.



[2]Broockman, David E. 2014. "Do female politicians empower women to vote or run for office? A regression discontinuity approach." *Electoral Studies* 34: 190-204.

Question 3 (Total: 12 points)

Angrist and Krueger (1992) use instrumental variable estimation to understand the effect of years of education on wages.[3] These researchers exploit two American education laws: (1) You can only enter school (first grade) when you are 6 years old by September 1st, and (2) you must remain in school until age 16.

To help us out, let $T = \begin{cases} 1 \text{ if born Jan-Aug} \\ 0 \text{ if born Sept-Dec} \end{cases}$

In sum, these laws mean that children born earlier in the year are older when they start school and therefore receive less schooling by the time they reach the legal drop-out age. For example, children born in 2000 can drop out of school on their birthday in 2016. Those born in January-August will be in 10th grade, while those born in September-December will be in 9th grade. Since a child's birth date is unrelated to intrinsic ability, these researchers argue it is a good instrument for years of education to identify the true effect of education on wages.

Recall the two equations we solve at the same time for instrumental variables estimation:

$x_i = \tau + T_i\gamma + \epsilon_{i1}$

$y_i = \alpha + x_i\beta + \epsilon_{i2}$

Answer the following questions.

(a) (3 points) What are T, x, and y?

(b) (3 points) What does β estimate?

[3] Angrist, Joshua D., and Alan B. Krueger. "The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples." Journal of the American statistical Association 87.418 (1992): 328-336.

(c) (3 points) What must we assume for instrumental variables estimation to correctly estimate β?

(d) (3 points) Do you think this study meets those assumptions? Why or why not?