Homework for Chapter 20: Regression Discontinuity

*How Does It Work?*

1. Regression discontinuity designs rely on a treatment being assigned based on whether the \_\_\_\_\_\_\_ variable is on one side or another of a \_\_\_\_\_\_\_\_. Just to one side, someone is much more likely to be treated than on the other side. In these cases, the effect of treatment can be estimated by estimating how much the outcome variable changes from one side to the other, within a narrow \_\_\_\_\_\_\_\_ of the cutoff.
2. Which of the following is the most accurate intuitive explanation of why a sharp regression discontinuity design can identify of a causal effect of treatment?
   1. Because when treatment is assigned with a cutoff, there are naturally no back doors between treatment and outcome.
   2. Because in a sharp regression discontinuity design, treatment is assigned by only one variable.
   3. Because it is reasonable to assume that falling just barely to one side of a cutoff or another is basically random.
3. Explain why it gives regression discontinuity designs a hard time when the variable that determines treatment is measured *granularly* – for example, measuring only which bucket someone falls into, when each bucket covers a range, rather than their actual value of the variable that determines treatment.
4. You are interested in the effect of a factory being inspected on whether it reduces its pollution output. There’s a variable “number of reported violations” that you have. When the number of reported violations goes from 9 to 10, the probability of being inspected jumps from 20% to 60%. Average pollution levels decrease from 600k metric tons of CO2 to 500k metric tons goingg from 9 to 10 as well. What is the fuzzy regression discontinuity estimate of the effect of inspections on air pollution levels?
5. The goal of estimating a regression discontinuity design estimate is to predict the average outcome just to either side of the cutoff. Why do we need to predict this? Why not just measure the outcomes for the observations that are closest to the cutoffs?

*How is it Performed?*

1. You are investigating the effect of being eligible for food assistance on your kids’ academic performance. You have data on income and childrens’ average test scores. Assume everyone becomes eligible for food assistance if their income is below $25,000 per year. You take all the data and estimate the effect by OLS, regressing test scores on (Income – 25000), Below (an indicator for income being below $25,000 per year), and (Income – 25000) x Below. You get the below regression table.

|  |  |
| --- | --- |
|  | **Test Scores** |
| (Intercept) | 74.999\*\*\* |
|  | (0.022) |
| (Income - 25000) | 0.001\*\*\* |
|  | (0.000) |
| Below | -4.978\*\*\* |
|  | (0.031) |
| (Income - 25000) × Below | 0.000 |
|  | (0.000) |
| Num.Obs. | 16132 |
| R2 | 0.996 |
| \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01 | |

* 1. What is the regression discontinuity estimate of the effect of food assistance on test scores?
  2. Why did we subtract 25000 from Income before running the model?
  3. Assume that all the important regression discontinuity assumptions about the research design hold (no cutoff manipulation, no other treatments at the cutoff, etc.). What is one major concern you’d likely have about the way you *estimated* this effect?

1. Some regression discontinuity estimates use a *kernel weight* like the triangular kernel, that doesn’t just apply a bandwidth (dropping observations too far away from the cutoff) but also weights observations more heavily. Why might this be a good idea, as opposed to just using a narrower bandwidth?
   1. Because it reduces the impact of observations that are likely to introduce more bias, but doesn’t drop them fully so as to increase the effective sample size and improve precision.
   2. Because it has the effect of de-biasing the impact of observations that are further away from the cutoff; weighting these observations closer to zero will eliminate the impact of bias before it eliminates the true effect, retaining only what we want.
   3. Because it is necessary to use local regression: the shape of the local regression in the limited bandwidth is the same as the shape of the bandwidth (linear in the triangular case).
   4. It is only a good idea in circumstances where we can expect that the shape of the relationship on either side of the cutoff is fairly linear and so the weighting can be used to include further-away observations rather than a fitted shape.
2. You are interested in the effect of library fines on whether long-overdue library books get returned. You find a local library who decided to forgive all library fines owed by people who owed below $50 in total, and want to use regression discontinuity. However, you talk to the library people and they say that a tiny handful of library patrons were insistent that they felt honor-bound to repay their fines, and wanted them reinstated. Also, a tiny handful of patrons with fines in the $51-55 range called the library many times until they got theirs removed too. So a few people below the cutoff didn’t get treated, and a few people above it did. If you estimated this using a sharp regression discontinuity estimator anyway, how would this affect your results and why? Assume the true effect of canceling fines is that it increases return rates by 10%.
3. One common placebo test applied in regression discontinuity is to rerun the regression discontinuity model, but use a different variable as the outcome. Generally, the different variable is one that we might consider to be on a back door between treatment and outcome. If we find a large effect of treatment on this variable, this will make us concerned about our regression discontinuity model. Why would we hope that this effect would be zero?
   1. Because if it’s nonzero that means that treatment is affecting things other than the outcome, and we have assumed that treatment only affects the outcome.
   2. Because we’ve assumed that being just on one side of the cutoff or the other is basically random, and if it is indeed related to something else, that assumption is likely wrong.
   3. Because we’ve assumed that, in the absence of treatment, our predictions on both sides of the cutoff would be the same (“smooth at the cutoff”). A nonzero effect shows that the outcome would have jumped at the cutoff.
   4. Because we’ve assumed that we have the shape of the relationship between outcome and running variable correct (linear, polynomial, etc.). A nonzero effect shows that we have the wrong functional form and need to try another.
4. You are analyzing the effect of two-factor authentication on having your identity stolen. At a certain bank, people are required to turn on two-factor authentication if they have $1,000 or more in their account. However, you look at the distribution of the amount of money in people’s accounts and find that there are far more people just below $1,000 than just above it. Explain intuitively why this should make you worried about your estimate.

Coding (which includes any how-the-pros-do-it questions)