

## PROBLEM SET 8: REGRESSION DISCONTINUITY

### MGMT 737

1. PART I (Comparing Estimators) We will replicate the main results from Table 1 of Huh and Reif (2021, AER Insights). This estimation will require using linear regression and also canned packages from two sources: Cattaneo et al.'s **rdrobust** software <https://rdpackages.github.io/rdrobust/> and Michal Kolesar's **rdhonest** package <https://github.com/kolesarm/RDHonest>.

The code and data for this paper is available here: <https://github.com/reifjulian/driving>

- (a) Replicate the estimated point estimate and confidence intervals for Column 2 for the **All causes** outcome. (Note the Readme file on the github page – it is quite easy to run this)
- (b) Now, re-estimate this using a uniform kernel. How does it change the estimates?
- (c) Re-estimate this using a local quadratic function. How does it change the estimates?
- (d) Finally, re-estimate this using a bandwidth of 40. How does it change the estimate?
- (e) Describe what the “covs” option in the estimation approach is doing. How does the main estimate change if you do not include it?
- (f) Re-estimate the main specification again, but drop observations where `firstmonth = 1`. How does this compare?
- (g) Now, re-estimate this RD, dropping observations where `firstmonth = 1`, and using the **RDHonest** approach with an  $M$  smoothness parameter of 0.1 and the bandwidth specified by the Cattaneo estimator. How do the estimates and the confidence intervals compare?
- (h) How small does the smoothness parameter need to be to match the confidence intervals from the original result? (approximately). How plausible does that seem?
- (i) How does your RD Honest estimate change if you increase the bandwidth to 40?
- (j) Finally, run the following two estimates:
  - i. **rdrobust** omitting observations where `firstmonth = 1` and a uniform kernel
  - ii. a linear regression of

$$\text{cod\_any} = \gamma_0 + \text{agemo\_mda}\gamma_1 + 1(\text{agemo\_mda} > 0)\gamma_2 + 1(\text{agemo\_mda} > 0) \times \text{agemo\_mda}\gamma_3 + \epsilon_{it} \quad (1)$$

using the observations that are within the bandwidth picked by the **rdrobust** command (hint this should be under 11, and is called `h`) and omitting observations where `firstmonth = 1`.

How does your point estimate of  $\gamma_2$  compare to the **rdrobust** estimate? Explain.

- (k) Explain, in words, how you could get the **rdrobust** estimate that uses the triangular kernel using OLS (hint: it would involve weights)