Problem Set 3
ECON 672 Winter 2022/2023
Due 1/26/2023

Timely compliance is an important component of regulatory agencies. The Office of Labor-Management Standards (OLMS) is the regulatory agency that fulfills the Labor-Management Reporting and Disclosure Act of 1959 (LMRDA), which requires annual financial files from every private-sector union. Unions are required to file their financial reports 90 days after their fiscal year ends (which typically is Dec 31, Mar 31, Jun 30, or Sep 30). https://www.dol.gov/agencies/olms/reports/forms/lm-1-lm-2-lm-3-lm-4. If the union files their financial report after 90 days, then they are not in compliance with the LMRDA. Unions with $250,000 in receipts need to file an LM-2 form, while unions with less than $250,000 in recipients need to file an LM-3 form or LM-4 form (for unions with less than $10,000 in receipts).

In 2005, OLMS had a policy change that required unions with 250K in receipts were required to file their LM-2 electronically instead by paper, while LM-3 and LM-4 forms could still be filed by paper until 2017. (See: https://www.dol.gov/agencies/olms/laws/regulations/nprm-required-electronic-filing-lm-3-ml-4 and https://apwu.org/news/department-labor-changes-rules-annual-lm-reports). Did this policy change improve timely compliance with the LMRDA? (Hint: the receipts data are skewed to the right, so try to focus on receipts less than $500,000).

The OLMS LM data have been limited to 2005 to 2016, which is the time period required for LM-2 forms to be filed electronically. The data can be found on GitHub or ELMS. Your outcome of interest is a binary "late_filing", which is 0 if they filed on time and 1 if they filed late. Your treatment of interest is "D", which is the cutoff for unions needing to electronically submit LM-2 forms.

Question 1: What would be the running variable X_i and what would be the cutoff c_0 [3 points]

Question 2: Estimate the OLS estimate with a biased treatment. What does it show? [3 points]

$$Y_i = \alpha + \delta D_i + \varepsilon_i$$

Question 3: Recenter the running variable around the cutoff and estimate a global linear regression along the running variable [3 points]

$$Y_i = \alpha + \delta D_i + \beta \tilde{X}_i + \varepsilon_i, \; where \; \tilde{X}_i = X - c_0$$

Question 4: Set a window around the cutoff and re-estimate the linear regression along the running variable [3 points]

You are free to set the window as you see fit

Question 5: Functional form of the running variable locally (window) and globally (all data) [6 points]

Question 5a: Estimate a quadratic functional form:

$$Y_i = \alpha + \delta D_i + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \varepsilon_i$$

Question 5b: Estimate a linear function form with interaction

$$Y_i = \alpha + \delta D_i + \beta_1 \tilde{X}_i + \beta_2 D_i \tilde{X}_i + \varepsilon_i$$

Question 5c: Estimate a quadratic functional form with interaction

$$Y_i = \alpha + \delta D_i + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \beta_3 D_i \tilde{X}_i + \beta_4 D_i \tilde{X}_i^2 + \varepsilon_i$$

Question 6: Use a cmogram function. What is a more appropriate functional form? What is the LATE estimate for the more appropriate functional form? [3 points]

Question 7: Use a local polynomial kernel function. What is the LATE estimate? Narrow the dataset to 2010 or later. What is the LATE estimate? [6 points]

Question 8: Use RDRobust. What is the LATE estimate? [3 points]

Question 9: Indirectly test the continuity assumption. [12 points]
Question 9a: Utilize a McCrary Test to test for continuous density along the running variable for all year. Do we fail to reject the null hypothesis of continuous density along the running variable at the cutoff?

Question 9b: Utilize a McCrary Test to test for continuous density along the running variable for 2010 and later. Do we fail to reject the null hypothesis of continuous density along the running variable at the cutoff

Question 9c: Test for any discontinuous jumps for liabilities, assets, or members at the cutoff. Are there any discontinuous jumps in these covariates at the cutoff?

Question 10: What is a potential problem or challenge to identification with using RDD for this population that we discussed in class which can be applied to the LM financial data? [2 points]