Problem Set 2
Econ 672
Due August 6, 2022

Question 1: We want to assess the returns to marriage to see if there is a premium for men similar to Cornwell and Rupert (1997) and a penalty for women. There is self-selection into marriage, so we need to develop an estimator to try to identify the effect of marriage on earnings. First, we will develop 1-year panels from the Current Population Survey (CPS). Then, we will need to assess a pooled OLS and a fixed effects estimator. If we assume that there is no time-varying heterogeneity (strict exogeneity assumption), what are the returns to marriage?

Please use the CPS data from 2020 and 2021 found in GitHub (https://github.com/rowesamuel/ECON672/tree/main/Data/Introduction) or ELMS. We will use a Mincer equation ($Y_i = f(education, experience, experience^2, treatment, covariates)$ to assess the marriage premium or penalty. Our outcome of interest will be the natural log of wage which will come from taking the natural log of "pternwa" in the CPS. Our treatment variable of interest will come from "pemaritl" in the CPS. This is a category variable that needs to be dichotomized into a binary variable (0 is not married and 1 is married). You will also need to develop potential experience from age (which is "prtage" in the CPS) and education from "peeduca" in the CPS. Please see the data dictionary for the CPS here: https://www2.census.gov/programs-surveys/cps/datasets/2022/basic/2022_Basic_CPS_Public_Use_Record_Layout_plus_IO_Code_list.txt

1) Append the monthly CPS files for 2020 and 2021 and develop a 1-year panel for individuals using the between 2020 and 2021. The panel should be strongly balanced with no gaps. The 1st interview should be in 2020 and the 8th interview should be in 2021. (Hint: use "hryear4 instead of hrmis for the time period in xtset.) [3 points]
2) What are the returns to marriage for men and women? Estimate a pooled OLS regression and estimate a Fixed Effects (Within Estimator). What is the difference between the two estimators? [3 points]
3) What are the returns to marriage for men or is there a marriage premium for men? What does the pooled OLS show? What does the Fixed Effects Estimator (Within Estimator) show? [3 points]
4) What are the returns to marriage for women or is there a marriage penalty for women? What does the pooled OLS show? What does the Fixed Effects Estimator (Within Estimator) show? [3 points]
5) Using a Fixed Effects estimator, can we interact race/ethnicity data with the marriage to estimate group level returns to marriage? Why or why not? [3 points]

Question 2: 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 (https://github.com/rowesamuel/ECON672/blob/main/Problem%20Sets/Problem%20Set%202/olms_late_filings.dta) 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.

1) What would be the running variable X_i and what would be the cutoff X_c? [3 points]
2) Estimate a simple regression analysis and show the biased estimate [3 points]
    a. $Y_i = \alpha + \delta D_i + \varepsilon_i$ show?
3) Recenter the running variable and estimate a linear regression globally along the running variable [3 points]
    a. e.g.: $Y_i = \alpha + \delta D_i + \beta \tilde{X}_i + \varepsilon_i$ where $\tilde{X} = (X_i - X_c)$
4) Set a window around the cutoff and re-estimate the linear regression along the running variable [3 points]
5) Functional Form and find the LATE [6 points]
    a. Estimate a quadratic function $Y_i = \alpha + \delta D_i + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \varepsilon_i$
    b. Estimate a linear interaction $Y_i = \alpha + \delta D_i + \beta_1 \tilde{X}_i + \beta_2 D_i \tilde{X} + \varepsilon_i$
    c. Estimate a quadratic interaction $Y_i = \alpha + \delta D_i + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \beta_3 D_i \tilde{X} + \beta_4 D_i \tilde{X}_i^2 + \varepsilon_i$
6) Use the cmogram function [3 points]
    a. What is the more appropriate functional form? What is the estimated LATE for the more appropriate functional form?
7) Use a local polynomial kernel function [6 points]
    a. What is the estimated LATE?
    b. Narrow the years to 2010 or later, what is the estimated LATE?
8) Estimate a RDRobust estimate [6 points]
    a. Include all years, what is the estimated LATE?
    b. Narrow the years to 2010 or later, what is the estimated LATE?
9) Testing Continuity Assumption [10 points]
    a. Utilize a McCrary Test to test for continuous density along the cutoff for all years. Do we fail to reject the null of continuous density along the running variable?
    b. Utilize a McCrary Test to test for continuous density along the cutoff for 2010 and later. Do we fail to reject the null of continuous density along the running variable?

c. Test to see if there are any discontinuous jumps in assets, liabilities, or members at the cutoff. Are there any discontinuous jumps in the covariates at the cutoff?

10) What are possible problems or challenges to identification with RDD that we discussed in class which can be applied to the LM financial data? [2 points]

Question 3: We want to assess the impact of "Right-to-Work" (RTW) laws on unionization. Our research question is the following: do RTW laws reduce the unionization? Looking at aggregate data, we are unable to assess individual movements in and out of unionization, so we will use individual panels from the NBER CPS MORG (cps_morg_2009_2016.dta) to assess the research question. RTW laws eliminate union security agreements or the requirement to pay union dues or agency fees to the union that has a collective bargaining agreement with a firm. In February 2012 Indiana adopted a RTW law, while Michigan implemented a similar law in late 2013 and Wisconsin in March 2015. You need to assess the Average Treatment on the Treated (ATT) for the adoption of RTW laws on unionization.

Your outcome of interest is union membership, which is binary 0 or 1. You need to set up a canonical 2-by-2 Difference-in-Differences to assess the ATT of RTW on union status. $Y_{its} = \alpha + \gamma Post_t + \beta Treat + \delta(Post * Treat)_{st} + \epsilon_{its}$. You will be utilizing a linear probability model (LPM), since the outcome is binary. Your treatment variables will be the following for each individual 2-by-2 DD: INTREAT, MITREAT, and WITREAT. Your post variable will be the following: INPOST, MIPOST, and WIPOST. These variables are 0 for the first observation (before the law) and 1 in the second observation (after the law) is passed or implemented.

You will need two qualifiers for each 2-by-2 DD regression. First, each state has a set of comparison states. For example, you can see Indiana's comparison states by tabulating "in_borderstate" and stfips. Second, you will need to observe the time period of interest. These have been set up for you. For example, Indiana's observations before and after its law for the same individual can be found with "inob1". As a final note, these data have 1-year panels for each individual with "ID" being the unit variable and "year" as the time variable. Use do not need xtreg, but you will need areg and absorbe(ID) to account for individual fixed effects that will take care of the taste hypothesis. Please answer the following questions:

1) Assess the ATT by running separate 2-by-2 DD [9 points]
    a. What is the ATT for Indiana's law, use inob1 as the time period?
    b. What is the ATT for Michigan's law, use miob1 as the time period?
    c. What is the ATT for Wisconsin's law, use wiob1 time period?
2) Specification tests [10 points]
    a. Adding covariates to each 2-by-2 DD (consider a Mincer equation), what are the ATT estimates?
    b. Adding individual fixed effects (Hint areg y x, absorb(ID)) for each 2-by-2 DD, what are the ATT estimates ?
    c. Are the results relatively robust when adding covariates? Do the coefficients on the three ATTs radically change or are they stable?

3) Placebo tests with a triple difference-in-difference-in-differences [15 points]
    a. Use “management” for individuals in managerial occupations as a within-treatment-state comparison group for a Triple DDD. An RTW law should not affect managerial workers. (Hint: use “i.management”)
    b. What is the parameter of interest for the DDD for each DDD estimated? Are they statistically significant? If so, does this valid or invalidate our DDs?
    c. Rerun the Triple DDD for each states, but include a private sector covariate in your 2-by-2 DD called “private” (Hint: use “i.private”). Do your results change? Are you DDD parameters of interest statistically significant?
4) Placebo/Falsification Tests [9 points]
    a. Rerun the 2-by-2 DD for each state, but use Jan 2010 as the false cutoff, which are inob2 for Indiana, miob2 for Michigan, and wiob2 for Wisconsin.
    b. Do you find any statistical significant results? Does this corroborate or invalidate the original results robust?
5) Standard Errors [2 points]
    a. Even if you clustered standard errors at the treatment group level, what is a potential problem with standard errors?