Problem Set 2 Econ 797B: Empirical Methods in Labor Economics

October 7, 2020

Due: October 27, 2020

1) A PDF with write-up of your answers, including tables and figures
2) A single .do file, or R code, with clearly delineated sections for each problem
3) Output from running the code (e.g., log file)

Problem 1: Minimum Wage, Family Income and Poverty

Does minimum wage policy contribute to fighiting poverty by raising the bottom wage? Or does it, as some economists have argued, reduce job availability among the least privileged and hence reduce family income at the bottom? In this problem set, you will use a small extract of the Census from 1990 and the American Community Survey from 2006 to answer these questions using a variety of tools taught in the class.

Dataset: **pointonepctsampleE.dta** - this is a dataset that draws a 1/2 percent sample of individuals (age 16-65) from each state each year with family income under the median income-to-needs ratio.

Key variables: minwage (the maximum of state or federal minimum wage), year, statefips (state FIPS code), division (code for the 9 census divisions) age, sex, married (marital status dummy), racesingd (single digit race code), citizen, hieduc (years of education), empstatd(detailed employment status), uhrswork (usual hours worked per week), inctot (total personal income), incwelfr (welfare or public assistance income), poverty (ratio of family income to poverty-level income based on family size: this variable takes on 0-500, where 100 is at poverty, 500 is income that's 5 times the poverty level, where it's topcoded). The family income measure includes all sources of cash income, including public assistance and earnings. *For ease of computation, I have cut off the sample at the median of the poverty variable (which in 1990 was about poverty=300, and in 2007 was poverty=331.)*

PartA (25 points)

Define three binary outcomes $I_j = 1(poverty < c_j)$ for c_j=50, 75, 100, 125, 150. For each estimate the following:

i) a difference in difference regression using:

> **outcome:** I_j
>
> **treatment:** log minimum wage
>
> Place and time fixed effects you need for a basic DID, but no additional covariates.

ii) same as (i) but with covariates you think are relevant for controlling for demographic and economic condition. (Hint: what kind of controls may be "bad?" Use all other ones listed in "key variables" besides "bad controls" and justify the exclusion of the "bad controls"

iii) instead of common time effects, use division-specific effects with (i) and (ii)

Produce a "journal style" single table - where columns are different specification, and have three horizontal panels (for the three cutoffs - use of esttab or estout is recommended.). Have "Notes" at the bottom of the table explaining what you are doing. In the writeup, discuss differences in identifying assumptions and interpretations of the results.

Part B (25 points)

i) Using the "division-specific time effect" and covariates, do distributional regressions for "poverty" cutoffs between 0 and 250 in increments of 25 (that's 10 regressions), and plot the coefficients and confidence bands using appropriate standard errors (picture should look similar to the Havnes/Mogstad or Dube (2019) papers, e.g., Figure 2 in Dube). What is the interpretation of the results?

ii) Now use these coefficients to estimate unconditaionl quantile partial effects (as in Firpo Fortin and Lemieux paper). In particular, divide the estimated coefficients from (i) by -1 $\times$ the value of the PDF of *poverty* at those cutoff values (i.e., 25, 50, etc.) Confirm for 2 of these points that you get the same coefficient as using the "rifreg" package in Nicole Fortin's website. Plot these values, and interpret them.

iii) **OPTIONAL**: take a few cutoff points (say poverty=50 and 100) and construct correct standard errors accounting for the estimation of the PDF using block-bootstrapping.

Problem 2: Synthetic Controls and Difference and Differences Case Study - Card (1992) Effect of California Minimum Wage

In July 1988, California's minimum wage rose from $3.35 to $4.25. Card (1992, http://davidcard.berkeley.edu/papers/minwage unemp.pdf) estimated the impact of this single case by comparing it to a specific set of control states: Arizona, Florida, Georgia, New Mexico, and part of Texas. In particular he looked at the effect of the policy on wages and employment of several groups, including teens. In this problem set we'll assess those findings using more modern approaches to identification and inference.

Dataset **emp_wage_data.dta**: State-by-quarter dataset 1980q1-2000q4. **But for our analysis we will focus on the 1982q1-1990q1 period.**

Key variables: teen_logwage(log teen wage), teen_emp (teen employment), MW (minimum wage), *_share* (demographic group shares by age, race, gender, married), *_ind* (employment share of 1 digit naics industries)

Part A

We want to study the impact on labor market outcomes for log teen employment and log teen wages. First find the potential control group ("donors") as those states that did not raise their minimum wages between 1982q1 and 1990q1 (there was a federal increase in 1990q2).

i) The approach in Part A is to use a basic difference-in-differences between CA and all donor states. Show the following outcomes in for all estimates. Log wages and log employment for teens, log wages and log employment for overall workforce. Show figures with the difference between CA and control states over time using quarterly bins. How do CA and other control states compare in terms of pre-existing trends? Make a table showing the overall difference-in-difference estimates for all outcomes.

ii) How sensistive are the estimates to using controls for pre-existing linear and quadratic trends (while being careful not to allow the trends to soak up any lagged effect of the treatment itself)?

iii) Now implement a propensity-score-reweighted version of the difference-in-differences method. Use pre-treatment period averaged covariates (age, race, hispanic shares) for the logit propensity score model. Repeat steps (i) with this method. What states get more weight? [Do this manually, but also use `teffects ipw` if using Stata as a check; should get similar estimates.]

iv) Since this is a single treated unit, the cluster-robust variance estimator are not consistent. To assess this, use Conley-Taber's method and the control states to calculate the distribution of the treatment effect $\hat{\beta}$ under the null hypothesis of zero effects. How does your infernce using the C-T method compare to using the cluster-robust method?

Part B

Construct a synthetic "synthetic California" using the method of Abadie et al. (STATA: "synth.ado"). The donor pool includes all other states that did not raise minimum wages during the sample period. You can try two synthetic control methods:

- All pre-treatment quarterly outcomes
- Averaged pre-treatment outcomes, and also pre-treatment period averaged:
  - industry shares of employment
  - demographic shares (race, hispanic origin, education [HSL], age, gender)

i) What states are picked by the synthetic control method to constitute "synthetic CA?" Show the weights by state name. (OPTIONAL: show a map using maptile function in STATA or some other method).

ii) Show the outcomes in "CA" versus "synthetic CA" over the 1983q1-1990q1 period. (HINT: use "keep(...)" option to store the dataset.) How well does the synthetic CA match actual CA? What is the estimated effect for the outcomes, and how do these compare with your estiamtes from part A?

iii) Loop over the donor states and perform the same analysis to create the cutoffs at the 90% level using randomization inference and a figure similar to the Abadie paper. HINT: For each iteration, allow all other states without a minimum wage increase to be in the donor pool. Save results in a dataset for each iteration, along with the original CA estimates. Merge these together by date, and then plot all the treat-synth outcome differences in one graph, making CA dark and the other 24 light grey like in the paper.

iv) What happens if you estimate the synthetic control model using 1983q1-1986q2 only, and then use 1986q3-1988q2 as a validation and the 1988q3-1990q1 for the post period?

v) Repeat the exercise in (iv) but now using Arkhangelsky et al. (2019) Synthetic DID, where you use the synthetic CA weights for donors, but do a difference in difference estimate. How do the estimates compare, both in the validation and the post period?

vi) OPTIONAL: test for robustness of synthetic CA weights to different uses of lagged outcomes and other controls. How sensitive are the weights is it to reasonable alterations of the controls? How well do different controls predict outcomes out of sample (in the placebo synthetic controls - like in Dube and Zipperer 2015)?