Pueden trabajar en grupos de hasta 5 personas. Pongan claro el nombre de todas las personas que trabajaron en el grupo. Suerte!

ECONOMETRÍA APLICADA AVANZADA

PROBLEM SET – Differences-in-Differences

Profesor(a): Cristina Tello-Trillo
Email: tellotri@gmail.com
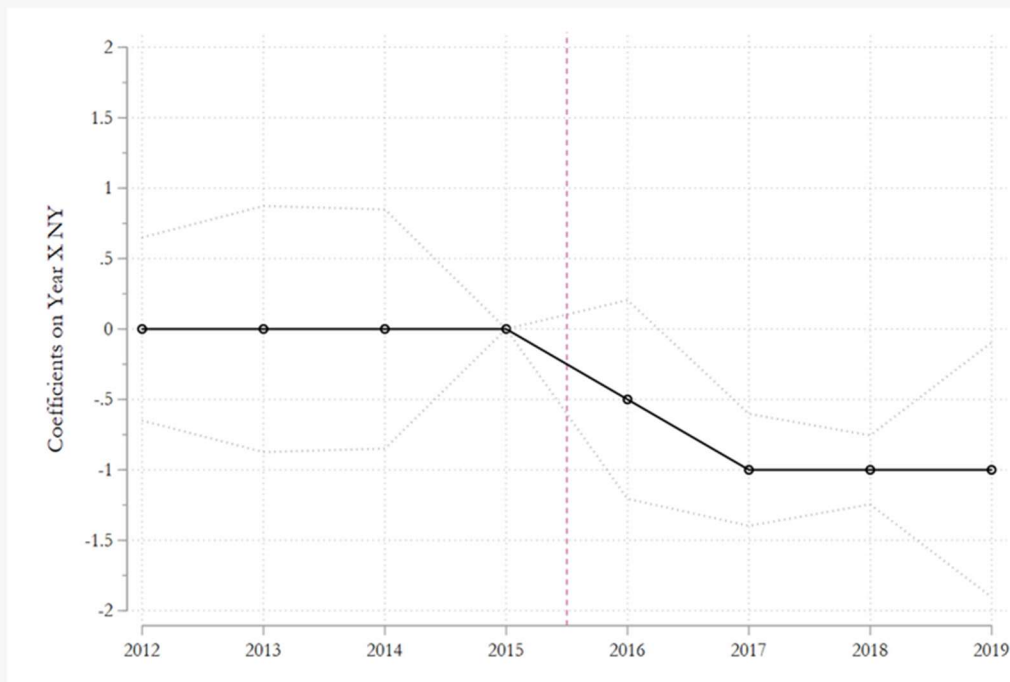
Problem 1 [20 points]

NY implemento una politica de "Paid Family Leave" (Descanso familiar pagado) en el 2016, y queremos evaluar como afecto los salarios de las mujeres. Se implementa un DiD en donde se compara NY con estados vecinos que no han implementado la política. Se estima el event study de la siguiente manera
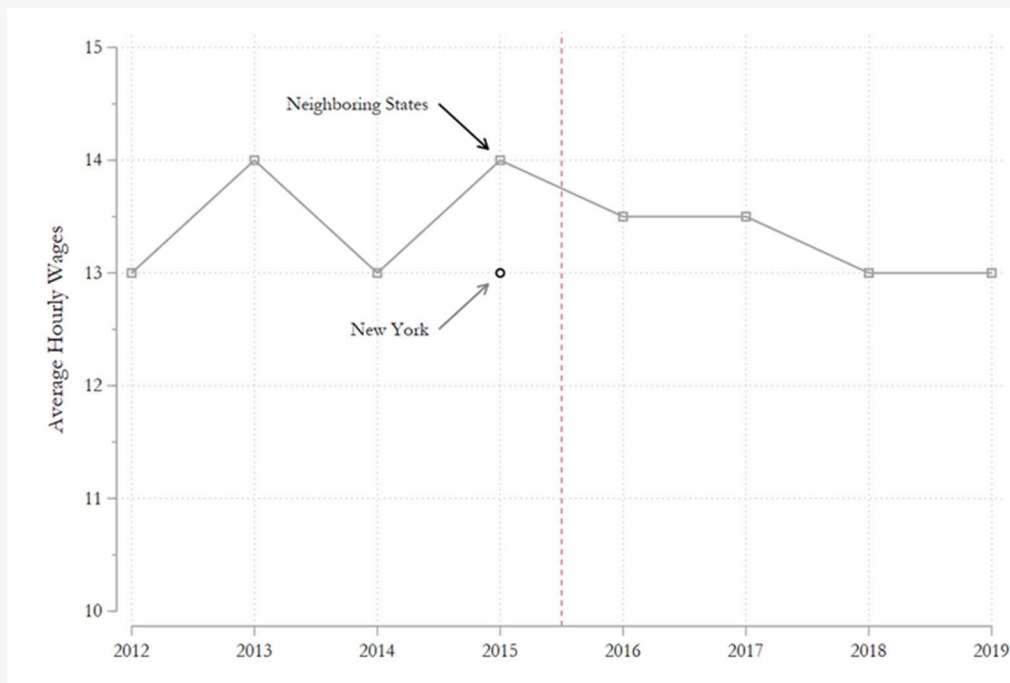
$$\text{Hourly Wages}_{ist} = \beta_o + \beta_{2012} NY \times 2012 + \ldots + \beta_{2019} NY \times 2019 + S_s + Y_y + \epsilon_{ist}$$

El año de base es 2015, un año antes de la reforma. El modelo include state and year fixed-effects. La figura 1 representa los coeficientes de la dummy de NY y las dummies de cada año. Usa la informacion de la figura 1 para completar el grafico de la figura 2.

Figura 1: Event Study

Figura 2: Trends

Problem 2 [50 points]

In this question, you will analyze the effect of China entering the WTO in 2001 on U.S. Manufacturing employment. Several papers attempt to tackle this issue (See Autor, Dorn, Hanson, 2013), we are going to mimic these papers using a simple framework. First, download from the Census website the County Business Patterns (CBP) series. This series includes the number of establishments, employment during the week of March 12, first quarter payroll, and annual payroll. CBP datasets are downloadable files in comma-separated value (CSV) format. In addition, corresponding record layouts and reference files are available. https://www.census.gov/programs-surveys/cbp/data/datasets.html

The technical documentation for the dataset is here:

https://www2.census.gov/programs-surveys/cbp/technical-documentation/records-layouts/full-layout/state_layout.txt

See the attached do file which explains how to extract the dataset for year 1999 and keep all the 3-digit NAICS sector. You will have to do the same for years 1998-2006 and append all these files together. Create a variable called "year" that tracks the year of each dataset, this will be useful

when appending all the years together. Keep the variables FIPSTATE, YEAR, NAICS, EMP, QP1, AP, and EST. Hint: Use the command "foreach var in " " " to loop across years.

1. [2 points] What is the level of observation?
2. [2 points] Construct 1 dummy variable called "post_china" where post_china=1 for year>=2001 and 0 otherwise.
3. [2 points] Construct 1 dummy variable called "manuf" where manuf=1 for all the observations that start with naics code "3" – which is manufacturing - and 0 otherwise.
4. [4 points] Construct the values necessary to generate the difference-in-difference estimate (i.e. 2x2 Matrix) of the effect of China entering the WTO on employment (emp). Hint: Define clearly what is your treatment group vs control group and the intervention time. Interpret the results.
5. [5 points] Estimate a diff-in-diff regression and make sure you get the same diff-in-diff estimate as in part 4.
6. [10 points] Estimate a diff-in-diff regression for the effect of China entering the WTO in 2001 on the number of establishment (est), an average pay (qp1/emp). Interpret the results.
7. [5 points] Estimate same regression as in (5) but now take logs of the dependent variable (i,e, log(emp)). Interpret your results. Is it necessary to take logs?

Now you will proceed to run an event study. You will analyze the trajectory of the effect of the China shock on different outcomes, employment, number of establishments, and average pay.

8. [6 points] Generate one dummy per year. Construct the interaction between each year dummies and your treatment group (manuf). You should have 9 interaction terms.
9. [10 points] Estimate an event study, i.e. run the following specification: log(emp) vs year dummies, manuf*year dummies (omit the interaction between manuf * year 1998) and control for NAICS-3 digit dummies and state dummies. Hint: when controlling for naics3 and state dummies, you need to use the command "reghdfe y x, absorb(naics3 state)." This will include naics3 dummies and state dummies in your regression. Interpret your results. Should you expect to see any effect for the interaction term manuf*year 1999 or manuf * year 2000? Did the China shock have a significant effect on employment? Was it a short-run or long-run effect?
10. [4 points] Estimate a similar event study on the log(est) and average pay. Interpret your results.

Problem 3 [30 points]

This exercise is based on the paper, Eissa, Nada, and Jeffrey B. Liebman. 1996. Labor Supply Responses to the Earned Income Tax Credit (EITC). Quarterly Journal of Economics. 111(2): 605-637. The data "eitc.dta" file contains data for single women 20-54 with less than high school education, as this group is most likely to be affected by the EITC.

1. [3 points] Create a table summarizing all the data provided in the data set.
2. [5 points] Calculate the sample means of all variables for (a) single women with no children, (b) single women with 1 child, and (c) single women with 2+ children.
3. [5 points] Construct a variable for the "treatment" called anykids (indicator for 1 or more kids) and a variable for time being after the expansion (called post93—should be 1 for 1994 and later).
4. [10 points] Using the "interaction term" diff-in-diff specification, run a regression to estimate the difference-in-differences estimate of the effect of the EITC program on earnings. Use all women with children as the treatment group.
5. [7 points] Repeat (iv), but now include state and year fixed effects [Hint: state fixed effects, are included when we include a dummy variable for each state]. Do you get similar estimated treatment effects compared to (iv)?
6. [7 points] Using the specification from (v), re-estimate this model including urate nonwhite age ed unearn, as well as state and year FEs as controls. Do you get similar estimated treatment effects compared to (v)?
7. [7 points] Estimate a version of your model that allows the treatment effect to vary by those with 1 or 2+ children. Include all other variables as in (vi). Does the intervention seem to be more effective for one of these groups over the other? Why might this be the case in the real world?
8. [6 points] Estimate a "placebo" treatment model as follows: Take data from only the pre-reform period (up to and including 1993). Drop the rest, or restrict your model to run only if year <= 1993. Estimate the effect for all affected women together, just as in (vi). Introduce a placebo policy that begins in 1992 (so 1992 and 1993 are both "treated" with this fake policy). What do you find? Are your results "reassuring"?