**Assignment 5: Due May 29, 2020.**

**Weiyi Li**

**05/27/2020**

**Casual Inference**

**Github link for other files used: https://github.com/weiyili96/assignment-5.git**

Two-way fixed effects

1. Fill in the missing information in the following table
2. Run the following regression in Stata or R and report the coefficient and standard error on .

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Unit** | **Time** | **Y** | **D** | **Mean of Y** | **Mean of D** | **Demeaned Y** | **Demeaned D** |
| 1 | 1 | 60 | 1 | 46.55 | 0.5 | 13.45 | 0.5 |
| 1 | 2 | 35 | 0 | 46.55 | 0.5 | -11.55 | -0.5 |
| 2 | 1 | 10 | 1 | 46.55 | 0.5 | -36.55 | 0.5 |
| 2 | 2 | 156 | 0 | 46.55 | 0.5 | 109.45 | -0.5 |
| 3 | 1 | 13 | 0 | 46.55 | 0.5 | -33.55 | -0.5 |
| 3 | 2 | 65.5 | 1 | 46.55 | 0.5 | 18.95 | 0.5 |
| 4 | 1 | 10 | 0 | 46.55 | 0.5 | -36.55 | -0.5 |
| 4 | 2 | 20 | 1 | 46.55 | 0.5 | -26.55 | 0.5 |
| 5 | 1 | 35 | 0 | 46.55 | 0.5 | -11.55 | -0.5 |
| 5 | 2 | 61 | 1 | 46.55 | 0.5 | 14.45 | 0.5 |

Where is a unit fixed effects and are time fixed effects.

1. Run the following regression and report the coefficient and standard error on .

How does your answer in (3) on differ from what you found in (2)? What does this tell you about the algebraic equivalence between an OLS regression that includes unit and time fixed effects and an OLS regression using the demeaned values of the outcome and treatment variable (controlling for time fixed effects)?

The 2 coefficients are the same. Because when we demeaned the values, it eliminates the unit fixed effect, then the expected value of estimated coefficient would be equal in these 2 equations.

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
| VARIABLES | y | demeanedy |
|  |  |  |
| d | -15.50 |  |
|  | (28.42) |  |
| unit | -6.700 |  |
|  | (9.846) |  |
| 2.time | 45 | 45.00 |
|  | (28.42) | (27.31) |
| demeanedd |  | -15.50 |
|  |  | (27.31) |
| Constant | 51.90 | -22.50 |
|  | (37.28) | (19.12) |
|  |  |  |
| Observations | 10 | 10 |
| R-squared | 0.335 | 0.284 |

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Difference-in-differences

1. The following dataset (simulation.dta) is a panel of 50 states, 25 firms per state, over 40 years yielding a dataset equaling 50,000 rows. There are two outcomes – y and y2. They were generated according to the following formula.

Where TE is a unique random **positively value** treatment variable associated with groups 1 to 5, TREAT is a dummy variable equaling 0 or 1 depending on if the units are treated in a particular year, and treatdate is the year in which a group was treated. Summarize the value of TE across the sample. What should the effect ( of the treatment on the outcome under constant treatment effects?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Obs | Mean | Std. Dev | Min | Max |
| te | 30,000 | 5.399668 | 2.260098 | 1.847116 | 10.1371 |

The effect of treatment is positive; which should be around 5.3997 under constant treatment effect.

1. Now run the following regression

Report the coefficient on TREAT. Is the coefficient on TREAT unbiased? How do you know?

The coefficient on treat is 5.295. This coefficient should not be biased, because from question 4 we know the effect should be positive on outcome; and the error range is really small there. And the result is statistically significant as well.

1. Now calculate the following group-time ATT. Group 4’s TE is equal to 2. Each subsequent year is 4 x number of years since treated.[[1]](#footnote-1) What is group 4’s treatment date? What should the treatment effect be for group 4 in 2006? In 2007? Should the treatment effect for all post-treatment periods be positive, negative or zero?

The treatment date for group 4 is 2004.

ATT (2004,2004) =4

ATT (2004,2006) =8

ATT (2004,2007) =10

The treatment for all post-treatment periods should be positive.

1. Now calculate the following regression. Assume the variance weighted common trends equals zero.

Report the coefficient on TREAT. Is the coefficient on TREAT unbiased? Why is it biased? Write down the formula from Goodman-Bacon (2018,2019) that explains what the OLS coefficient equals with dynamic treatment effects. If the VWATT equals your answer in (6), and the VWCT equals zero, then what must be the size of the heterogeneity bias with dynamic treatment effects?

|  |  |
| --- | --- |
| treat | -21.82\*\*\* |
|  | (0.913) |

This coefficient is biased because given on what we have known, the treatment effect should be positive for all post-treatment periods. But under this estimation, the effect is negative.

δ = VWATT + VWCT – ΔATT

If the VWATT equals what I had in (6), and VWCT is 0; this means that it attenuates the estimate bias and can even reverse sign be depending on the magnitudes of what is otherwise effects in the sign in a reinforcing direction.

1. The Callaway and Sant’anna (2019) (“CS”) estimator is a nonparametric estimator using propensity scores weighting with a “not yet treated” group as the control for comparison on the long difference. Write out the estimator below.

ATT(g,t) = E [ (Gg/E[Gg] - ( 1-Dt)/E[1-Dt] ) \* ( Yt - Y\_{g-1} ) ]

1. Going back to question (6). Use the CS estimator to calculate the group-time ATT for group 4 in 2006 and 2007. Is this unbiased?

ATT (2004,2006) =9.181719

ATT (2004,2007) =9.979418

This is **unbiased** in this example, because we have a true control group and have been treated for group 4.

But if in some other examples that there’s no true control group and never treated, then this might have a problem because it’s calculating 2\*2s by comparing later treated units to early treated units, which shouldn’t be correct.

1. See lines 90 and 98 and 117. [↑](#footnote-ref-1)