

Problem Set 5

Instructions: Answer the following questions and submit your results via email to `sean.corcoran@vanderbilt.edu`. Use your name and problem set number as the filename. The Stata log for Q1 should include the instructions, your commands, and output. Graphical output may be submitted separately, preferably as a PDF file. Working together is encouraged, but all submitted work should be that of the individual student.

Question 1. This problem will use regression difference-in-differences to estimate the impact of a breakfast in the classroom (BIC) program on school meals program participation in New York City. BIC was not implemented under random assignment; rather, schools voluntarily adopted the program. We do, however, have data for these and other schools before and after adoption. (See Corcoran, Elbel, & Schwartz 2015 for details). **(40 points)**

- (a) In Stata, open the panel dataset called *NYCbfastlunch-v2.dta* from Github:

```
use https://github.com/spcorcor18/LP0-8852/raw/main/data/NYCbkfastlunch-v2.dta, clear
```

This file consists of school-level data in which the rows are elementary or middle schools observed in year t ($t=2005$ to 2012). The outcome variables of interest are *bkfast_part* and *lunch_part*, which are average daily participation rates in the school breakfast and lunch programs. Provide some descriptive statistics for these two variables. On what scale are they measured? **(2 points)**

- (b) Use `xtset` to declare the data as a panel. Which variable is the cross-sectional unit and which is the time dimension? Is this a balanced panel? (Show how you answer the last question). How many schools are observed in all 8 years? **(2 points)**
- (c) This dataset contains a time-varying treatment variable called *bicpost* that is equal to 1 in years in which the school offered BIC and 0 otherwise. For this part, limit the sample to schools that adopted BIC in 2010 and never-adopters. Estimate a simple “2x2” difference-in-differences regression comparing mean breakfast participation rates for two groups (treated and untreated schools) in two time periods (pre and post). Do the same for lunch participation rates, and interpret your results. (Note: ignore the panel nature of your data and just use the `regress` command for this part). Is the BIC effect statistically significant? Practically significant? What assumptions must be satisfied for this difference-in-differences to be considered a causal effect? **(6 points)**

- (d) Repeat part (c) but include year dummies in the regressions. How does this affect your impact estimates for BIC, if at all? Continue to use the `regress` command, and explain why one of the post-2010 year effects is not estimable. (**4 points**)
- (e) In parts (c)-(d) you did not take advantage of the panel nature of your data. Re-estimate these regressions using `xtreg` and school fixed effects. How do your estimates compare? (**4 points**)
- (f) Compare the mean baseline characteristics of treated and untreated schools in parts (c)-(e). (Use 2009 as the baseline year). Look at the following: total enrollment, % ELL, % special education, % Asian, % black, % Hispanic, % female, % free lunch eligible (*free1*), % reduced price lunch eligible (*redu1*). How do schools that adopted BIC in 2010 compare to those that never adopted? (**4 points**)
- (g) Now estimate the same regression models in part (e), but include the school covariates listed above in your regressions. How do your estimates of the “BIC effect” change, if at all? (And how are these covariates related to meal participation?) (**4 points**)
- (h) Repeat part (e)—the model without covariates—but try using the Stata command `xtddidregress`. Compare your results. (**4 points**)
- (i) State the assumption(s) that are required to hold in order to interpret the breakfast participation DD coefficient in part (h) as causal. Use the post-estimation commands `estat trendplots` and `estat granger` to probe these assumptions, and interpret the results. (**5 points**)
- (j) Finally, incorporate all treated schools into the analysis regardless of when they adopted BIC. Estimate the two-way fixed effects model using `xtreg` (as in part e) and `xtddidregress` (as in part h). Try these without and with school covariates. How do your results compare? (**5 points**)

Question 2. For these questions, refer to the recent article by Cellini and Turner (2019), “Gainfully Employed? Assessing the Employment and Earnings of For-Profit College Students Using Administrative Data.” You can find the article here: <http://jhr.uwpress.org/content/54/2/342.abstract>. **(32 points)**

- (a) Cellini and Turner use a generalized difference-in-differences regression model to estimate the causal effect of attending a for-profit certificate program on labor market outcomes. What specific outcome variables do they examine, and what dataset(s) do they use? **(4 points)**
- (b) How is the “treatment” variable defined here and what are the possible “pre” and “post” years? How many potential pre and post years are there? **(4 points)**
- (c) The authors use three different groups of “untreated” individuals as comparison groups. What were they, and what was their rationale for looking at each? Which comparison group is their “preferred” one, and why? **(4 points)**
- (d) Equation (1) on page 350 shows their regression specification. Carefully explain what each term represents, and how the causal effect of attending a for-profit certificate program is being identified. Why is there not a main effect for “For-Profit” in the model? **(5 points)**
- (e) Carefully explain the main assumption necessary to interpret the difference-in-differences estimate here as a causal effect. What evidence do the authors provide that this assumption holds for their three different comparison groups? **(5 points)**
- (f) The paper’s main results are reported in Table 3. Carefully interpret the coefficients reported in Panel B. What additional evidence does Figure 2 provide? **(5 points)**
- (g) Finally, Figure 4 shows the distribution of earnings effects *by school* for public and for-profit institutions. Cellini and Turner describe these as the result of “single-difference” regressions. Briefly explain what they mean by this, and why these should not be interpreted as the causal effects of attending specific institutions. **(5 points)**