Regression II
Vanderbilt University
Prof. Sean P. Corcoran

Due October 29, 2024
41 total points

Problem Set 6

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. Working together is encouraged, but all submitted work should be that of the individual student.

Question 1. This problem will use the dataset described in the midterm. You are interested in the effects that eliminating school fees had on childrens' school enrollment in sub-Saharan Africa. You have annual country-level data on school enrollment rates from 1981 to 2015 for 15 countries and the year in which each country eliminated fees. (Note the enrollment measure is "gross enrollment" which can be greater than 100% since it is the ratio of total enrollment in a grade level divided by the population of the age group typically served by that grade, multiplied by 100). The countries eliminated fees in various years between 1994 and 2012. (**41 points**)

(a) In Stata, open the panel dataset called *WDI-FPE-data.dta* from Github:

```
use https://github.com/spcorcor18/LPO-8852/raw/main/data/WDI-FPE-data.dta, clear
```

For this problem, drop all observations from 2006 and later. All countries were eventually treated in this panel, and we would like to have some "never treated" cases for sake of this problem. After doing this, create a table and an accompanying bar graph that shows the years treatment occurred for countries in this panel. (You may decide on the format for these). How many "timing groups" are there? How many "never treated" countries are there? (**5 points**)

(b) Use `xtset` to declare the data as a panel. Is this a balanced panel? (**2 points**)

(c) Estimate the effect of eliminating fees on (i) primary school enrollment, and (ii) secondary school enrollment using "generalized difference-in-differences." For each, implement the two-way fixed effects regression model via `xtreg` and `xtdidregress`. When using `xtreg` be sure to cluster your standard errors appropriately. Briefly summarize your results. (**5 points**)

(d) (Quietly) re-run the `xtdidregress` command in part (c)—for both outcomes—and follow each by the post-estimation command `estat bdecomp` which provides the Bacon decomposition. Include the `graph` and `summaryonly` options. Explain in words what the results tell us. (**5 points**)

(e) The Stata command `xthdidregress` was created to implement a variety of estimators that allow for heterogeneous treatment effects and address the biases of traditional TWFE estimators when there is staggered treatment timing. Estimate the effect of eliminating fees on primary school enrollment using (i) the Wooldridge (2021) modified TWFE estimator, and (ii) the Callaway and Sant'Anna regression adjustment estimator. These are estimated using `xthdidregress twfe` and `ra`, respectively. (It may help to refer to the Stata help and documentation). Note the default is to use the never-treated as the comparison group. Explain in words how to interpret the regression output (in general, not each specific estimate). (**6 points**)

(f) (Quietly) re-run the `xthdidregress ra` command in part (e) and follow it by the post-estimation commands below. Briefly explain what each does. Is there evidence of heterogeneous treatment effects by timing group? By event time? How do these results compare to the traditional TWFE estimates in part (c)? Based on the *dynamic* plot, in what post-treatment years is the treatment effect statistically significant? Finally, explain why the standard errors get larger in this plot over time. (**8 points**)

```
estat aggregation
estat atetplot
estat aggregation, dynamic graph
estat aggregation, cohort graph
```

(g) As an alternative approach, use the command `eventdd` to estimate an event study model for primary school enrollment. Create your *event time* variable carefully, and use the option `method(fe)` to include country fixed effects. Be sure to also include calendar year effects. In words, interpret the *lead4* and *lag4* coefficients. Does the event study plot support the parallel trends assumption? Briefly explain. (**6 points**)

(h) Modify your `eventdd` syntax in part (g) to include the `accum` option and to consolidate "8 or more" time periods before and "8 or more" time periods after into one parameter (each). (See the `eventdd` help menu). (**4 points**)