Data Task for Prospective Research Assistants

Background

As part of the *Medicare Modernization Act* (MMA) of 2003, a small number of hospitals were given an upward adjustment in the wage index used to determine their Medicare reimbursement rates. The net result of these adjustments was that recipient hospitals experienced a sudden positive spike in their per-case reimbursement rates, increasing the revenue they received for the Medicare cases they saw. What made this adjustment economically interesting was that these adjustments were offered to hospitals as political "sweeteners" to shore up support for the MMA amongst representatives in whose districts the hospitals were located. As a result, the adjustments made under the MMA formed a plausibly exogenous policy shock that allows us to examine how hospitals respond to a sudden increase in funding. In this task, you will investigate whether hospitals that received a reimbursement rate adjustment under the MMA (also known as a "Section 508 waiver") increased their adoption of medical technologies as a result of those adjustments.

Tasks

Please complete the following tasks. Your work in tasks 1-5 should be completed in Stata, and accompanied by a .do file that allows for reliable replication of your analysis. Tasks 3 & 6 ask for written responses to various questions which you can complete in a word processor of your choice. Please put all your answers to these questions in a single document and name the write up file using the following convention: '[yyyymmdd]_[lastname]_[firstname]_writeup' where [yyyymmdd] will be the date on which you submit the task. Any tables or graphs you create should be included in the same write up file and come after your written responses; each table or graph should be on its own separate page!

1. Merge the provided data files (**datatask_main.dta** and **datatask_treat.dta**) to join the variables on hospital technologies and characteristics with the treatment status indicator. The treatment status indicator is equal to 1 for hospitals that received a Section 508 waiver and 0 otherwise.

2. Using the formulas provided in the attached excerpt (i.e. datatask_saidin.pdf), construct a Saidin Index score for all hospitals in each year.

   Output: Name the variable you create in this step `saidin` and save it with your merged dataset to a separate file.

   Hint: Just apply the formulas given to the data. There are variations on the Saidin Index with further nuances, but you don't need to worry about any of those complications here.

3. Using the data provided, answer the following questions and provide supporting evidence in the form of a graph or table. Restrict your analysis to 2004.

   a) Describe how the distribution of Saidin Index scores varies across the sample of hospitals.

b) Looking at the Saidin Index scores you've calculated, do different types of hospitals have different rates of technology adoption?

c) Do hospitals with more beds tend to have higher Saidin index scores?

Output: Provide written answers to each question, with supporting evidence, in your write up. (Also make sure to include the code used to produce your results in your .do file.)

4. Estimate the effect of receiving a Section 508 waiver on the adoption of medical technologies using the following difference-in-difference model:

$$y_{it} = \alpha + \rho_i + \lambda_t + \beta_t D_i + \epsilon_{it}$$

Where y_{it} is the Saidin Index for hospital i in year t, ρ_i and λ_t are fixed effects for the hospital and year, respectively, D_i is a treatment status dummy, and ϵ_{it} is a hospital-year specific error term. β_t is the treatment effect specific to year t, and is allowed to vary over time. Note that the treatment is "on" for treated hospitals in all years of the panel. Please cluster your standard errors at the hospital-level.

Output: Create a separate Stata dataset called "[lastname]_[firstname]_estimates.dta". This dataset should have 10 observations (i.e. one for every year) and it should contain the following 6 variables:
1) **year**
2) **tr_effect** - your estimates for β_t
3) **tr_lo** - the lower bound estimate of β_t using a 95% confidence interval
4) **tr_hi** - the upper bound estimate of β_t using a 95% confidence interval
5) **cr_mean** - your estimates for λ_t
6) **tr_mean** - $\beta_t + \lambda_t$

Please store all 6 variables as numeric (i.e. do not convert them to strings).

Hints: 1) When you estimate this model, please let 2004, the year of treatment, be the base ("omitted") category. Framed different, all point estimates should be normalized to zero in 2004. Think about what the coefficients estimated for the periods before and after 2004 represent to see why this is. 2) Manually entering the values you estimate into a new dataset creates room for mistakes. There are many ways to automate this process, but you may find the –parmest– package especially useful.

5. Use the estimates you produced in task 4 to create a single, cleanly formatted graph presenting your results. For the sake of comparison, your graph should have two lines showing the average change in the Saidin Index relative to 2004 for the control and treatment groups (i.e. **cr_mean** and **tr_mean**).

Output: Export your graph from Stata and include it in your write up. (The code for creating and exporting your graph should be included in your .do file.)

Hint: An excellent graph would also show 95% confidence interval bounds for the treatment effect. This is a style of graph we use in many of our projects. For example, my recent work on politics and hospital behavior uses graphs just like this (without confidence intervals). My attached NBER Working Paper (w23748.pdf) might be a handy reference.

6. Answer the following the following questions in your write up, based on your analysis so far:

   a) What can you conclude about the impact of MMA reimbursement rate adjustments (the "treatment") on hospital technology adoption? How confident are you in those results?

   b) What are the threats to identification in the difference-in-differences model used? Are the assumptions necessary for identification under that model met in the case you examined?

   Output: Your written answers.

Included Files

1. **datatask_treat.dta**: Contains a dummy variable capturing the "treatment status" of each hospital in the sample.

2. **datatask_main.dta**: Contains variables describing hospital characteristics, as well as dummy variables for whether a hospital possessed a particular technology in a given year. (You don't need to worry about what each of these technologies are—they've already been restricted to only those technologies we want to include in our index.)

3. **datatask_saidin.pdf**: Excerpt from Baker & Spetz (1999)[1] detailing the construction of the Saidin Index. Assume for the purposes of this exercise that the sample provided includes all relevant data for the population of hospitals we want to use for the calculation of the index.

4. **w23748.pdf**: A copy of my NBER working paper on politics and hospital behavior (Cooper et al., 2017).

---

[1] Baker, Laurence, Joanne Spetz. "Managed Care and Medical Technology Growth." *Frontiers in Health Policy Research, vol. 2*. 1999. Pp. 31.