ECON 470 HW2

Ben Yang

2023-02-15

Instructions

In this assignment, you'll recreate the HCRIS data and answer a few questions along the way. The first step is to make sure you're working with the HCRIS GitHub repository and downloaded all of the raw data sources. Once you have the data downloaded and the code running, answer the following questions:

Summarize the data

1. How many hospitals filed more than one report in the same year? Show your answer as a line graph of the number of hospitals over time.

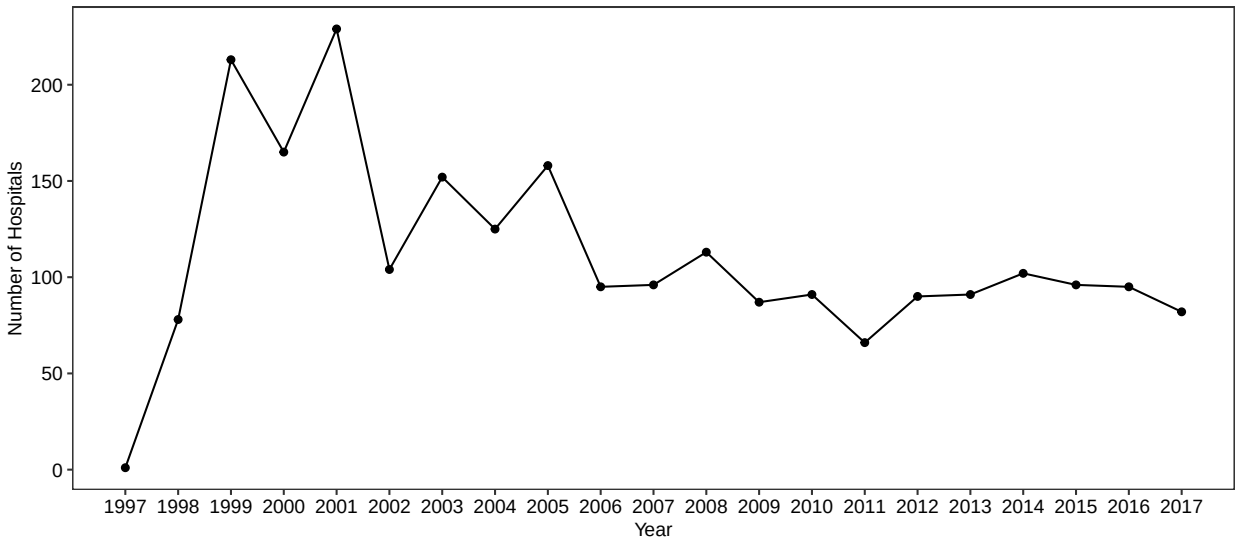

Figure 1: Number of Hospitals with More Than 1 Report in Each Year from 1997 to 2018

2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

After removing/combining multiple reports, there are 9323 unique hospital IDs (Medicare provider numbers) exist in the data.

3. What is the distribution of total charges (tot_charges in the data) in each year? Show your results with a "violin" plot, with charges on the y-axis and years on the x-axis. For a nice tutorial on violin plots, look at Violin Plots with ggplot2.

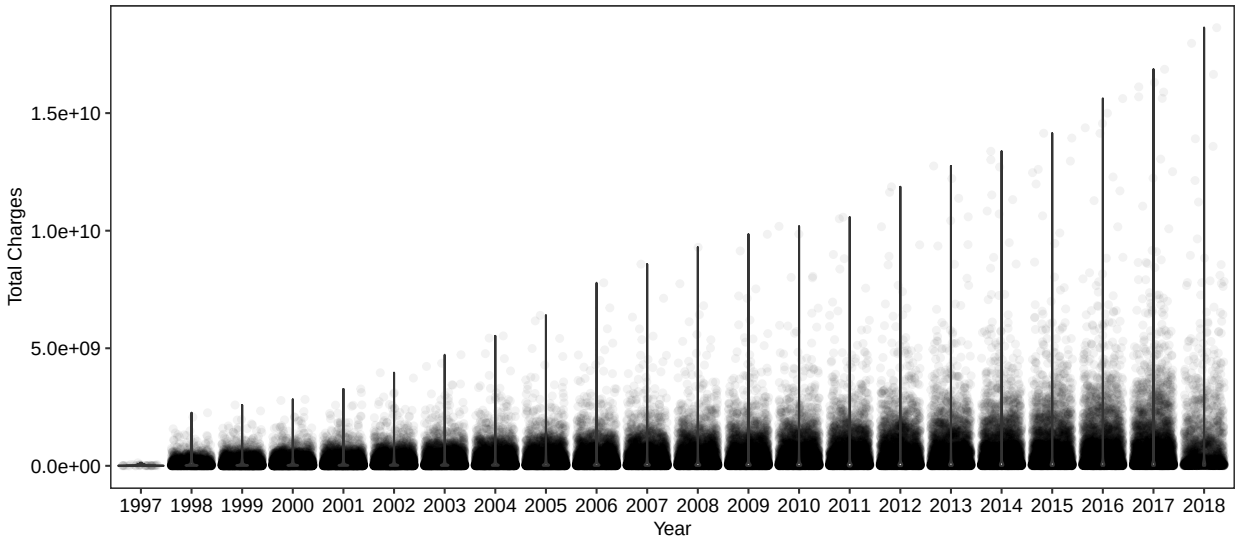

Figure 2: Distribution of Total Charges in Each Year from 1997 to 2018

4. What is the distribution of estimated prices in each year? Again present your results with a violin plot, and recall our formula for estimating prices from class.

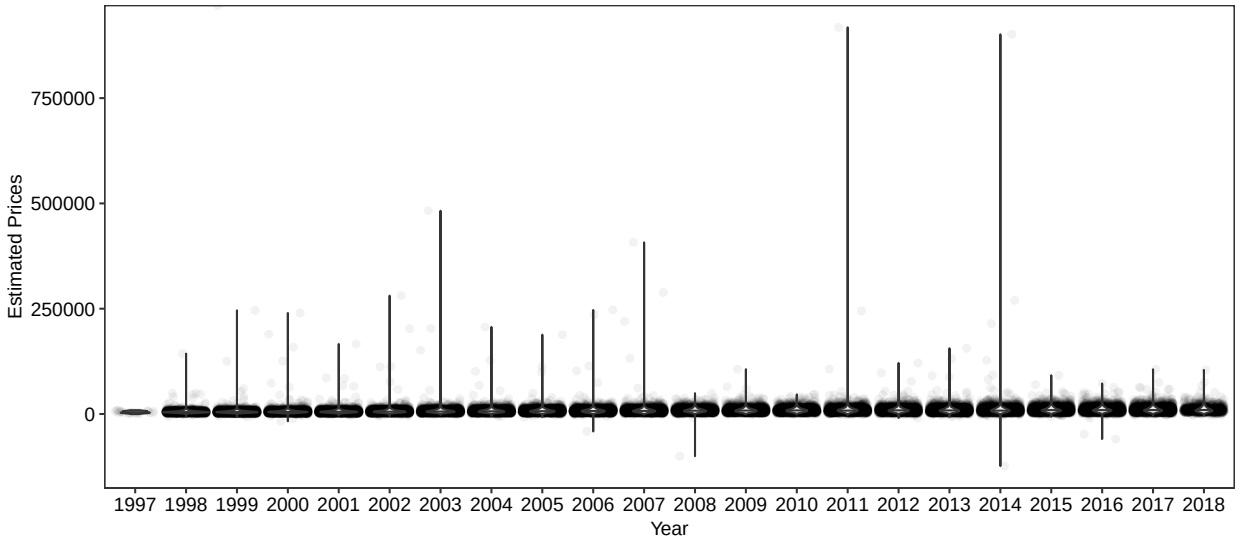

Figure 3: Distribution of Estimated Prices in Each Year from 1997 to 2018

Estimate ATEs

For the rest of the assignment, you should include only observations in 2012. So we are now dealing with cross-sectional data in which some hospitals are penalized and some are not. Please also define penalty as whether the sum of the HRRP and HVBP amounts are negative (i.e., a net penalty under the two programs). Code to do this is in the Section 2 slides.

5. Calculate the average price among penalized versus non-penalized hospitals.

The average price among penalized is 9896.3084978. The average price among non-penalized hospitals is 9560.4132266. The mean difference is 335.8952712.

6. Split hospitals into quartiles based on bed size. To do this, create 4 new indicator variables, where each variable is set to 1 if the hospital's bed size falls into the relevant quartile. Provide a table of the average price among treated/control groups for each quartile.

Table 1: Average Price among Treated/Control Groups for Each Quartile

penalty	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
Non-Penalized	7,678.757	8,507.619	9,869.173	12,329.72
Penalized	8,355.043	8,662.349	10,102.451	12,068.48

7. Find the average treatment effect using each of the following estimators, and present your results in a single table:

- Nearest neighbor matching (1-to-1) with inverse variance distance based on quartiles of bed size
- Nearest neighbor matching (1-to-1) with Mahalanobis distance based on quartiles of bed size
- Inverse propensity weighting, where the propensity scores are based on quartiles of bed size
- Simple linear regression, adjusting for quartiles of bed size using dummy variables and appropriate interactions as discussed in class

Table 2: Average Treatment Effects by Different Estimators for Each Quartile

	All	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile	Average
NN Inverse Variance Distance	-526.950000	-6.5365	-742.9100	-678.3100	-891.7900	-579.8866
NN Mahalanobis Distance	-492.820000	-175.8300	-803.1200	-716.9800	-834.5900	-632.6300
NN Propensity Score Distance	-201.030000	757.0600	-769.7500	-495.4400	-861.7000	-342.4575
Inverse Propensity Weighted Regression	-196.892217	571.7623	-705.2790	-438.2391	-918.5116	-372.5668
Simple Linear Regression1	-5.845761	482.5524	-700.4704	-423.2928	-1,019.7749	-415.2464
Simple Linear Regression2	-5.845761	482.5524	-700.4704	-423.2928	-1,019.7749	-415.2464

8. With these different treatment effect estimators, are the results similar, identical, very different?

The results of nearest neighbor matching (1-to-1) with inverse variance distance and with Mahalanobis distance are identical. The results of inverse propensity weighting and simple linear regression are similar. The results of the first two estimators and the last two estimators are very different.

9. Do you think you've estimated a causal effect of the penalty? Why or why not? (just a couple of sentences)

I think I have estimated a causal effect of the penalty, as the four aforementioned estimators suggest that the average treatment effect found in the data are statistically significant. The analysis considers various factors as the models examine the impact of penalty while controlling bed size, Medicaid Discharges, IP Charges, Medicare Discharges, and Total Medicare Payments.

10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated you.

Initially, I got stuck at question 6 as I was not able to create the appropriate quartiles of bed size for later analysis. Later I spent much time familiarizing with the regression codes. One thing I learned is that penalty may lead to higher hospital prices on average. What aggravated me is that the penalty may not lead to better services but only higher prices as the healthcare providers pass on the cost to patients.