Homework 2

Instructions

In this assignment, you'll recreate a more complete version of the Medicare Advantage data. The first step is to familiarize yourself with the Medicare Advantage GitHub repository and the associated data files.

The due date for initial submission is **2/10**, the revision due date is **2/12**, and the final due date is Friday, **2/13**.

Summarize the data

For our initial summary, we will work with data from 2014 to 2019. Based on these years of data, please address the following:

1. Remove all SNPs, 800-series plans, and prescription drug only plans (i.e., plans that do not offer Part C benefits). Provide a box and whisker plot showing the distribution of plan counts by county over time. Do you think that the number of plans is sufficient, too few, or too many?

2. Provide frequency histograms showing the distribution of plan bids in 2014 and 2018. How has this distribution changed over time?

3. Plot the average HHI over time from 2014 through 2019. How has the HHI changed over time?

4. Plot the average share of Medicare Advantage (relative to all Medicare eligibles) over time from 2014 through 2019. Has Medicare Advantage increased or decreased in popularity?

Estimate ATEs

For the rest of the assignment, you should include **only observations in 2018**. As we did in class, please define "competitive" markets as those with HHIs in the lower 33rd percentile of the national distribution of HHI, and define "concentrated" or "uncompetitive" markets as with with HHIs in the upper 66th percentile. This is somewhat arbitrary but it allows us to define a binary treatment variable in a way that we can more easily implement the methods in this module.

5. Calculate the average bid among competitive versus uncompetitive markets.

6. Split markets into quartiles based on Medicare fee-for-service (FFS) costs. To do this, create 4 new indicator variables, where each variable is set to 1 if the FFS costs falls into the relevant quartile. Provide a table of the average bid among treated/control groups for each quartile.

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

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

9. Pick your favorite flavor of estimators in this section (matching, weighting, regression, etc) and re-estimate treatment effects using the continuous FFS costs variable as well as total Medicare beneficiaries as your covariates. How does this result compare to the analogous estimate when matching/weighting only on FFS quartile?

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 or surprised you.