Homework 2

Research Methods, Spring 2026

Answer Key

My answers to the homework questions are described below. Note that I build the individual data for these answers in separate scripts. You can read in the full data as part of your markdown file, but that takes some time to compile to pdf. The GitHub repository for this work is available here. Enjoy!

Summarize the data

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?

The goal of this graph is to illustrate how many plans are available to an average enrollee in an average county. We remove SNPs, 800-series plans, and Part D only plans in order to focus on a more similar product that is available to everyone. For consistency with the rest of the analysis, I've also removed plans with missing plan IDs, plans in US territories, and plans not operating in an approved service area. The resulting box and whisker plot is provided in Figure 1, which provides a sense of the distribution of plan counts across counties in each year.
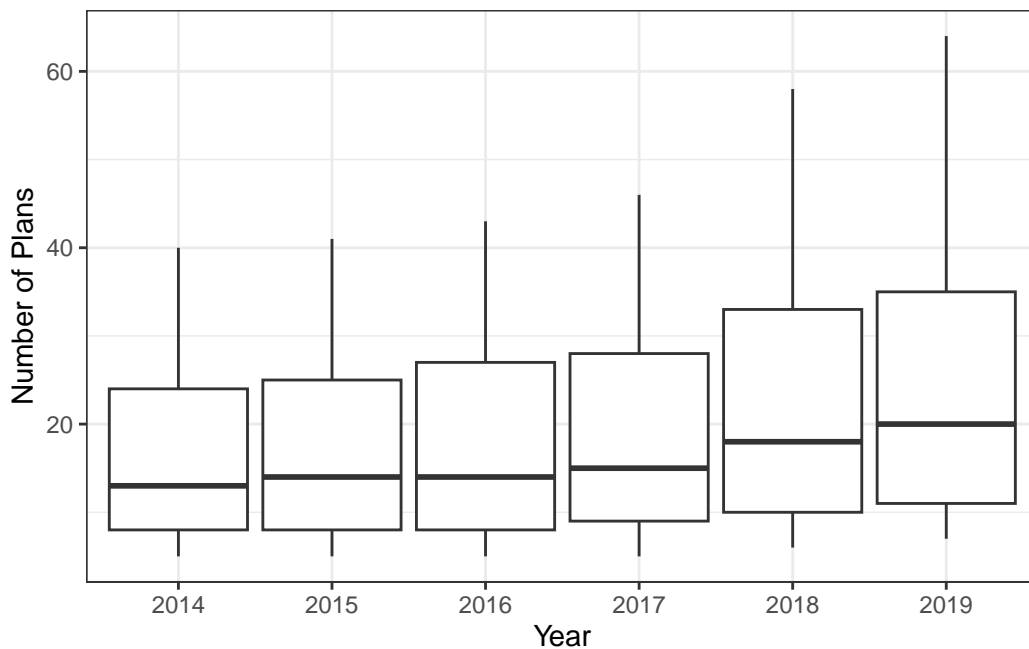

Figure 1: Plan Counts by County over Time

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

The distribution of bids is presented in Figure 2. As is evident from the figure, the distribution of bids largely overlaps between 2014 and 2018; however, there has been a slight leftward shift over time, reflecting a slight decline in MA bids over this time period (in nominal dollars). This is likely due to increased CMS pressure on MA plan bids. In particular, the ACA lowered and restructured MA county benchmarks, tying them more tightly to local FFS spending and reducing benchmarks most in high-spending areas. Since MA plans bid largely according to benchmark levels (seeking to keep bids below benchmark whenever possible), this induced plans to submit lower bids over time, with average bids falling from about 98% of FFS in 2014 to roughly 90% by 2018 (MedPAC 2014 and 2018 reports).
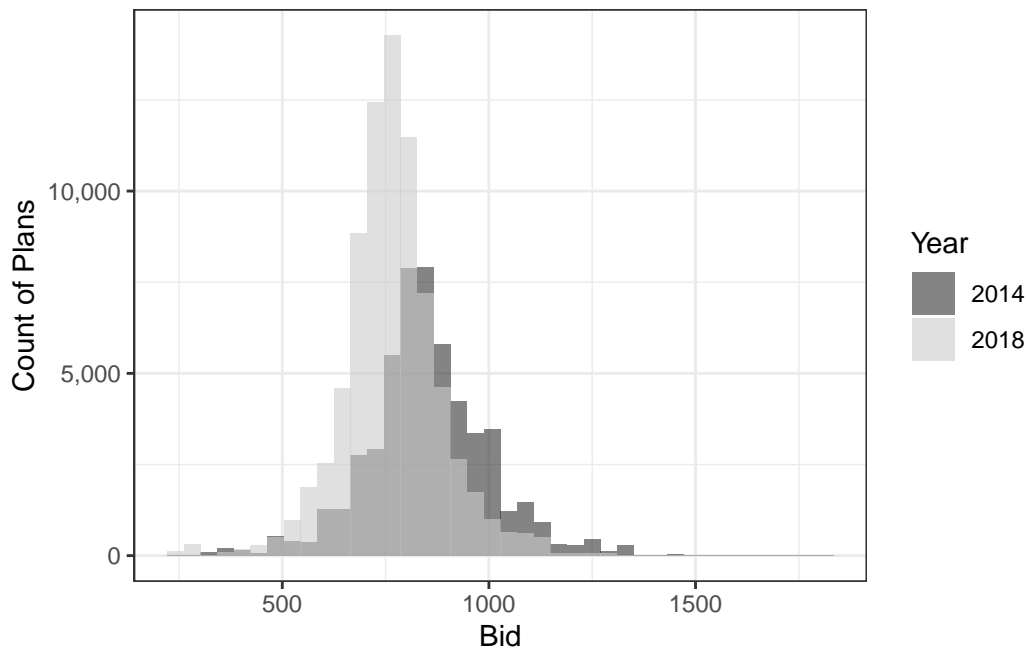

Figure 2: Bids for Selected Years

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

Average HHIs are presented in Figure 3. As we can see, HHI has actually been decreasing over time, indicating that MA markets have become more competitive over this time period. This is consistent with CMS and MedPAC reports indicating that MA plan participation has increased over this time period, with more plans entering markets and competing for enrollees. Note that, while local markets have become more competitive, the overall national market (at the insurer level) for MA plans remains fairly concentrated, with average HHIs above 2,500 throughout this time period.
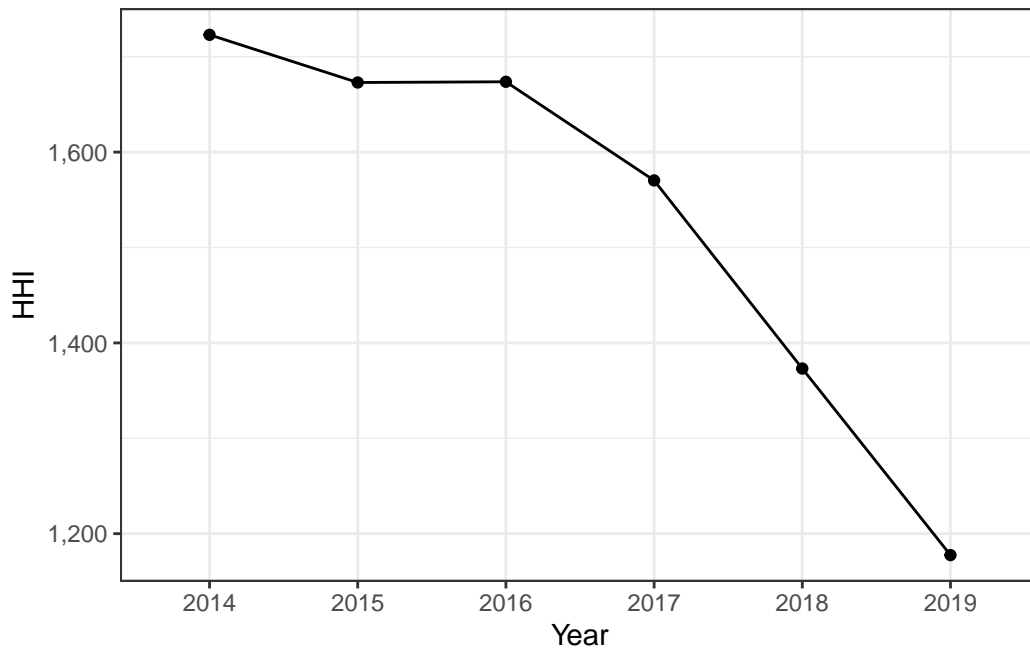

Figure 3: Average HHI over Time

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?

Average MA market shares are presented in Figure 4, showing a clear increase in MA enrollment over this time period. This is again consistent with CMS and MedPAC reports indicating that MA enrollment has been steadily increasing over this time period, reaching nearly 50% of all Medicare beneficiaries in recent years.
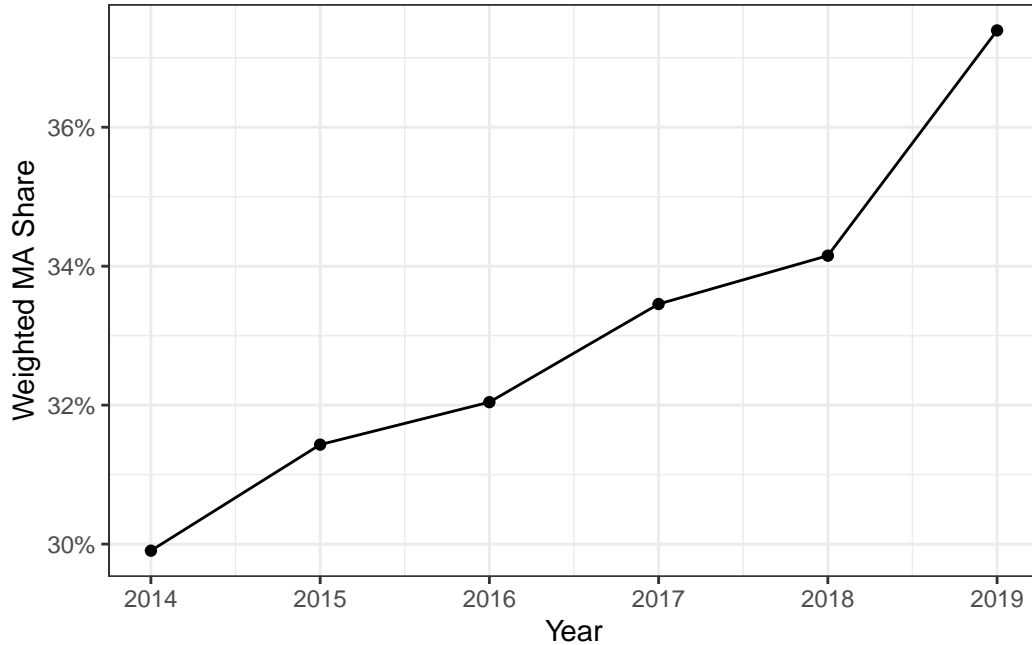

Figure 4: Average MA Share over Time

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.

Mean bids, as well as mean premiums, FFS costs, and eligibles, are presented in Table 1. We present results for both treated (high HHI) and control (low HHI) markets, as well as overall means for the same sample. As we can see, concentrated markets (high HHI) have slightly higher average bids and premiums than competitive markets (low HHI), as we might expect. Note that FFS costs and eligibles are also slightly higher in concentrated markets, which could be driving some of the differences in bids and premiums that we observe here.

Table 1: Balance table (2018)

	Low HHI	High HHI	Overall
Part C Premium	14.63	11.27	12.92
Plan Bid	782.33	787.03	784.69
FFS Cost	9,558.83	9,145.69	9,348.36
Medicare Eligibles	26,699.17	11,346.46	18,900.61

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. For this, you'll need to incorporate the FFS costs data as well.

Results are summarized below in Table 2, showing average bids by HHI level and FFS cost quartile. Surprisingly, bids are highest in the lowest FFS cost quartile; however, this could reflect many other differences across counties that are correlated with FFS costs. Nonetheless, we also see that higher HHI markets tend to have higher bids within each FFS cost quartile, consistent with the idea that less competitive markets lead to higher prices.

Table 2: Mean Bids by HHI and FFS Cost Quartile

FFS Quartile	Low HHI	High HHI
1	792.6511	798.7635
2	779.9991	784.5312
3	776.4584	775.1417
4	781.1339	787.4314

7. Find the average treatment effect using the estimators listed in the homework, and present your results in a single table.

Results are presented in Table 3. Note that all point estimates *should* be identical provided you are using the same data in each step, though the standard errors can differ. If your estimates differ across estimators, then it's probably because you have some missing observations somewhere that are dropped in one analysis and not in another.

Table 3: ATE Estimates

	INV	MAH	IPW	OLS
High HHI	3.896	3.896	3.896	3.896
	(2.606)	(2.606)	(2.601)	(2.600)

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

We've tried lots of different estimators, and in all cases, we've gotten the exact same answer! That's pretty cool and shows us how these estimators are all trying to do the same things. Note that the equivalence between these estimators is not true in general...it's only because we're considering dummy variables as our only covariates. If we had continuous variables as covariates, then these different estimators would be similar but not identical as they are here (which is what we found in class)

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?

Results are presented in Table 4 and suggest that what looked like a potentially positive effect of HHI on bids in Table 3 may have been driven in-part by size of the county (as measured by total Medicare eligibles) or the discretization of FFS costs into quartiles. Once we control for these factors more flexibly, the estimated effect of HHI on bids becomes smaller in magnitude and statistically indistinguishable from zero. This suggests that the relationship between market concentration and MA plan bids may be more tenuous than we initially thought, at least once we control for other important market characteristics.

Table 4: ATE Estimates

	IPW Baseline	IPW Alternate
High HHI	3.896	0.480
	(2.601)	(2.535)

10. Briefly describe your experience working with these data.

No wrong answers here. I'm just looking for you to reflect a little bit on this assignment.