

Exercise 2 Solutions

Ian McCarthy, Econ 771

Fall 2020

Overview

This document lays out my approach to addressing the questions in Exercise 2. There are lots of ways to answer these problems and build the data, especially in R! My way of answering the questions is just one of many.

I also have a particular workflow that you may or may not choose to adopt. That workflow is to do all of my analysis separately from my markdown document. I really don't like seeing huge markdown documents constantly going in and out of data work and discussion. It's also not feasible to work with really large data within the markdown document. So, I do all of the analysis, remove any large objects (including the data), keep only the relevant objects for the final markdown (figures, tables, specific numbers or statistics), and then save the workspace. Then I load the workspace into my markdown document.

Data

Before we get into the specific questions for this assignment, let's briefly go through the data. My code files for the benchmark data, penetration data, and AHRF data are included in the classroom repo under the filenames *1_benchmark.R*, *2_penetration.R*, and *3_ahrf.R*.

Benchmark data

MA benchmarks reflect a payment that CMS sets for each enrollee of a MA plan. The payment varied by county. CMS doesn't actually pay this rate to each plan - the final payment depends on the plan's "bid". This is not a "bid" in the standard auction sense. It's just the terminology used by CMS. Essentially, each plan sets their premium, and if the premium is below the benchmark rate, CMS pays the plan a risk-adjusted benchmark rate for each enrollee, and plans must use the

bid-benchmark differential to expand plan benefits. If the premium exceeds the benchmark rate, then CMS pays the plan the benchmark rate, and the plan passes the bid-benchmark differential onto the enrollee in the form of a monthly premium.

Beginning in 2012, benchmark payments also vary by the star rating of each MA plan. I didn't worry about this for our assignment. Instead, I just set the county-level benchmark rate equal to the benchmark rate for a 3-star plan, as follows:

```
mutate(bench_pay=
  case_when(
    year<2012 ~ risk_ab,
    year>=2012 & year<2015 ~ risk_star3,
    year>=2015 ~ risk_bonus0,
    TRUE ~ NA_real_
  ))
```

Penetration data

Enrollments for MA plans are published at the plan/county/year/month level. But in reality, plan enrollments don't change much across months in the same year. For the most part, enrollees are locked into whatever plan they choose for the year. There are exceptions, and anyone can drop a plan if they choose, but there are restrictions to which plans people can switch into outside of the open enrollment period. To make things easier, I just took the penetration data for December of each year for this exercise. In my full MA repo, there is some code to work with each month and average enrollments over the year if you are so inclined.

AHRF data

The AHRF data are easy to download but can be tricky to read into R. The best solution is to use the `SAScii` package, which allows you to read in the full ascii (.asc) file using the .sas input file provided as part of the data documentation. It is a slow process but works just fine. Then you just need to identify the variable of interest using the technical documents (*F15299* in our case).

Final dataset

Once you have the benchmark, penetration, and AHRF data together, you just have to merge them all. The easiest way to do this in our case is to merge on SSA and Year. Here's my code to merge the three datasets:

```
full.data <- ma.penetration.data %>%
  left_join(ma.benchmark.data %>% select(year, ssa, bench_pay, starts_with("risk")),
            by=c("ssa","year")) %>%
  left_join(ahrf.data %>% select(year, ssa, medicare_ffs_exp),
            by=c("ssa","year")) %>%
  filter(year>=2010,
         complete.cases(bench_pay, medicare_ffs_exp, penetration))
```

Note that I limit the data to “complete cases” (i.e., units of observation in which all of the main variables are observed in the data). This is a good rule-of-thumb for any empirical work because it ensures that you’re working with the same sample across different analyses; although it’s not always possible if you’re working with lots of outcomes.

Question and Answers

Question 1.

Provide and discuss a table of simple summary statistics showing the mean, standard deviation, min, and max of the Medicare Advantage penetration rate, benchmark payments, and utilization measures.

Table 1 presents the mean, standard deviation, min, and max of the relevant variables. Note the

Question 2.

Plot the mean county-level utilization against the mean MA penetration rate over time.

Figure 1 presents the MA penetration and Medicare expenditures, averaged across counties by year.

Question 3

Using the MA benchmark as an IV for MA penetration rates, estimate the effects of MA penetration on utilization. This is analogous to the specification in Baicker, Chernew, and Robbins (2013).

I’ll focus on a very simple specification here, using only the instrument (MA benchmark rates), the endogenous variable (MA penetration rate), and some county/year fixed effects. The syntax for this in R is:

```
felm(medicare_ffs_exp ~ 0 | year + ssa | (penetration ~ bench_pay) | ssa,  
      data=full.data)
```

The IV results are presented in Table 2.

Question 4

Present the “first stage” and “reduced form” results of your IV estimates in Part 3.

Quick refresher: The first stage is just the regression of the endogenous variable on all of the exogenous variables and your instrument, and the reduced form is a regression of your outcome on all the exogenous variables and the instrument. Recall that an IV estimate is just the ratio of the reduced form and the first stage, so it’s good to look at these things specifically. The results are presented in Table 3.

Question 5

Present selected specification/robustness checks of your IV and discuss two weaknesses of this identification strategy.

There’s no wrong answer here. I just want you to think critically about our identification and data. I’m going to check two issues:

1. Are the results sensitive to the use of the 3-star benchmark rate?

I replaced the benchmark payment rate with the 3.5-star benchmark payment rate just to see how the results change. Coefficient estimates are presented in Table 4.

2. Are the results sensitive to outliers in Medicare FFS expenditures?

Outliers are known to be a potentially major problem in any IV estimate. To examine the sensitivity of results to such outliers, I consider the “robust IV” estimator from Freue, Ortiz-Molina, and Zamar (2013). This is easily implemented with the R package that accompanies their paper, `riv`.

This “robust instrumental variables” estimator uses a “robust multivariate location and scatter S-estimator” and ultimately provides an IV estimate that is resilient to the presence of outliers (much more resilient than the standard IV). The “S-estimator” (I think) just denotes that we are “simultaneously” estimating both a location (measure of centrality) vector and scatter (measure of

dispersion), but we are doing so with a slightly different objective function other than the sum of squared errors.

In this case, you have to do a little extra work to fully replicate our original results. That's because we have lots of fixed effects...trying to include all of them explicitly as dummies in the RIV estimator will take forever. We can do this by taking advantage of the Frisch-Waugh-Lovell theorem and partialling out all of the fixed effects (year and ssa) from each variable of interest. In this case, we just estimate a linear regression of each variable on all of the fixed effects, take the residuals, and then run the RIV on the residualized values. Here's the code to do it:

```
y.resid <- lm(medicare_ffs_exp ~ factor(year) + factor(ssa),
             data=riv.data)
x.resid <- lm(penetration ~ factor(year) + factor(ssa),
             data=riv.data)
z.resid <- lm(bench_pay ~ factor(year) + factor(ssa),
             data=riv.data)

riv.data <- riv.data %>%
  spread_residuals(y.resid, x.resid, z.resid) %>%
  filter(complete.cases(y.resid, x.resid, z.resid))

Y <- as.matrix(full.data[,30])
Xend <- as.matrix(full.data[,31])
Zinst <- as.matrix(full.data[,32])

robust.iv1 <- riv(Y, Xend, Xex=NULL, Zinst, method="classical")
robust.iv2 <- riv(Y, Xend, Xex=NULL, Zinst)
```

Note that the RIV package is a little clunky. To accommodate the package, we need to drop incomplete cases (rows with any missing values). I've also explicitly defined the variables/covariates as their own matrices.

The coefficient estimate for MA penetration on Medicare payments using the default S -estimator is -86.48. Note that this is quite a bit larger than our original estimate of -46.164. Just to check, I also compared the “partialled-out” version of our estimator with a classical IV. This is the `robust.iv1` object in the above code. This should yield the same point estimate as the standard IV estimate earlier. Thankfully, they are the same.

Tables and Figures

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
MA Share	19.09	12.88	0.17	213.60
Benchmark Pymt	795.22	71.63	280.09	1,358.13
FFS Expenditure	9,747.87	1,169.40	4,876.11	20,451.58

Table 2: IV Results

	<i>Dependent variable:</i>
	Medicare Exp
MA Share	-46.164* (25.238)
Observations	18,352
R ²	0.785
Adjusted R ²	0.742
Residual Std. Error	593.073 (df = 15245)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 3: First Stage and Reduced Form IV Results

	<i>Dependent variable:</i>	
	MA Share	Medicare Exp
	(1)	(2)
Benchmark Pymt	−0.011*** (0.002)	0.510** (0.258)
Observations	18,352	18,352
R ²	0.947	0.797
Adjusted R ²	0.937	0.755
Residual Std. Error (df = 15245)	3.238	577.475
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Table 4: Alternative IV Results

	<i>Dependent variable:</i>		
	Medicare Exp	MA Share	Medicare Exp
	(1)	(2)	(3)
MA Share	-46.164* (25.238)		
Benchmark Pymt		-0.011*** (0.002)	0.510** (0.258)
Observations	18,352	18,352	18,352
R ²	0.785	0.947	0.797
Adjusted R ²	0.742	0.937	0.755
Residual Std. Error (df = 15245)	593.073	3.238	577.475
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01		

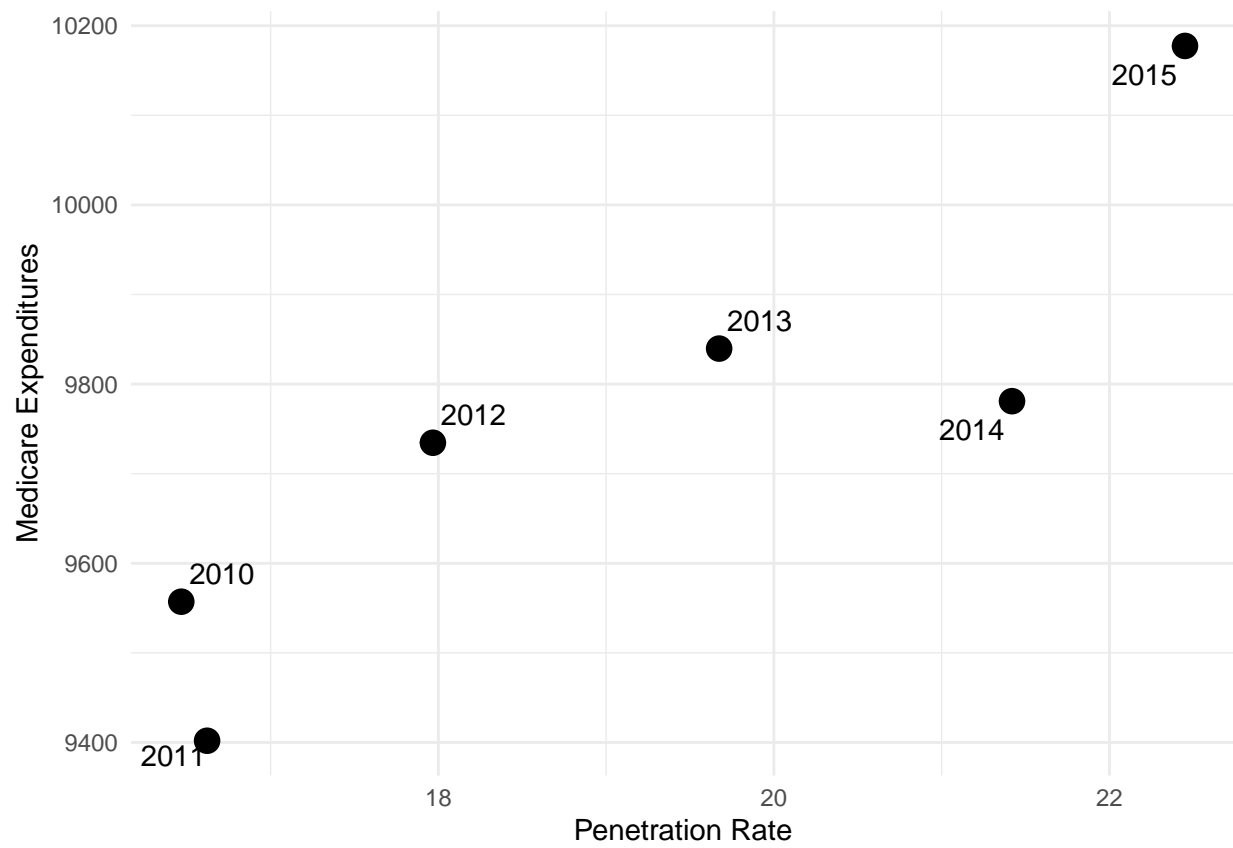


Figure 1: Mean MA penetration rates and payments by year

References

Freue, Gabriela V Cohen, Hernan Ortiz-Molina, and Ruben H Zamar. 2013. “A Natural Robustification of the Ordinary Instrumental Variables Estimator.” *Biometrics* 69 (3): 641–50.