Empirical Assignment 1

Katie Leinenbach

October 16, 2022

1 Summary Statistics

Below is the table describing summary statistics for physician-level medicare information. In 2013, average spending and claims drop and then steadily increase until 2017. Average patients seen drops between 2012 through 2017 while the minimum level of patients remains constant. Interestingly, the year that had the largest maximum medicare claims had the lowest minimum claim as well. 2016 had the largest average spending and claims even though it had one of the lowest average number of patients seen.

Total Physician-Level Medicare Spending

	Mean	Standard Deviation	Minimum	Maximum
2012	138424.5	271752	3.67	26288558
2013	136884.8	272486.7	5.94	20312469
2014	136262.6	271652	8.86	18764327
2015	137706.5	276931	0.93	17249304
2016	138787.2	289919	8.66	18345216
2017	138378.1	297068.4	15.35	22901522

Total Physician-Level Medicare Claims

	Mean	Standard Deviation	Minimum	Maximum
2012	2626.051	11317.26	11	2931819
2013	2550.189	10149.65	11	1122440
2014	2642.714	13019.41	4	3377317
2015	2770.554	15389.63	4.4	5750425
2016	2813.479	13665.8	5.2	1282350
2017	2788.634	13904.52	11	1428723

Total Physician-Level Medicare Patients

	Mean	Standard Deviation	Minimum	Maximum
2012	1046.699	2092.597	11	724713
2013	1026.01	1877.519	11	721303
2014	1006.201	1758.144	11	545218
2015	1011.889	1870.007	11	605766
2016	1006.842	1887.602	11	635666
2017	982.4349	1938.788	11	656227

Figure 1: Summary Statistics on Medicare Spending, Claims, and Patients

2 Mean Total Claims for Integrated vs. Non-Integrated Physicians

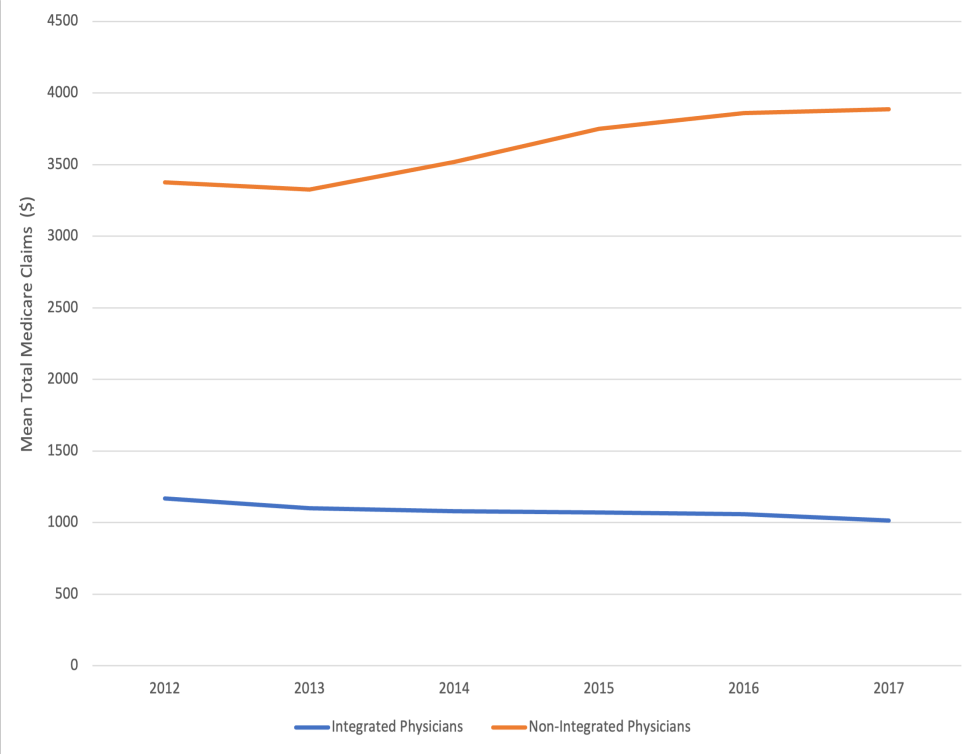

Figure 2: Mean Total Medicare Claims by Integration Status

3 OLS

Table 1: OLS Results

	Integration on Claims
Dependent Variable:	log(Total Claims)
Intercept:	-0.3223*** (0.0047)
Fixed Effects:	
NPI:	Yes
Year:	Yes
S.E. Clustered:	By NPI
Observations:	2,553,058
R2:	0.90194
Within R2:	0.00890

4 Estimated Treatment Effect Bounds

The table below shows the different estimated bounds for the treatment effect based on varying values of ρ and R^2_{max}. As ρ and R^2_{max} vary, the estimated right bound varies significantly. This indicates that there is a serious problem with the OLS estimation. These results support the need for an instrumental variable.

Table 2: Estimated Treatment Effect Bounds

Rho	**R Squared Max**	**Left Bound**	**Right Bound**
0	0.5	-0.3465131	-0.3465131
0	0.6	-0.3465131	-0.3465131
0	0.7	-0.3465131	-0.3465131
0	0.8	-0.3465131	-0.3465131
0	0.9	-0.3465131	-0.3465131
0	1	-0.3465131	-0.3465131
0.5	0.5	-0.3465131	1.3293314
0.5	0.6	-0.3465131	0.908344
0.5	0.7	-0.3465131	0.4873565
0.5	0.8	-0.3465131	0.0663691
0.5	0.9	-0.3465131	-0.3546183
0.5	1	-0.3465131	-0.7756058
1	0.5	-0.3465131	3.0051759
1	0.6	-0.3465131	2.163201
1	0.7	-0.3465131	1.3212261
1	0.8	-0.3465131	0.4792513
1	0.9	-0.3465131	-0.3627236
1	1	-0.3465131	-1.2046985
1.5	0.5	-0.3465131	4.6810204
1.5	0.6	-0.3465131	3.4180581
1.5	0.7	-0.3465131	2.1550958
1.5	0.8	-0.3465131	0.8921334
1.5	0.9	-0.3465131	-0.3708289
1.5	1	-0.3465131	-1.6337912
2	0.5	-0.3465131	6.3568649
2	0.6	-0.3465131	4.6729151
2	0.7	-0.3465131	2.9889654
2	0.8	-0.3465131	1.3050156
2	0.9	-0.3465131	-0.3789342
2	1	-0.3465131	-2.0628839

5 2SLS

Table 3: 2SLS Results - First Stage and Reduced Form

Dependent Variable:	Integration	log(Total Claims)
Practice Revenue Change	1.35e-5*** (6.56e-7)	
Intercept:		-3.688*** (0.1991)
Fixed Effects:		
NPI:	Yes	Yes
Year:	Yes	Yes
S.E. Clustered:	By NPI	By NPI
Observations:	2,252,137	2,252,137
R2:	0.89231	0.81312
Within R2:	0.00247	-1.0330

6 DWH Test

Below are the results from the DWH test when done manually. R also reports the values. I get the same results for both.

Table 4: DWS Results

Dependent Variable:	log(Total Claims)
Intercept:	-3.688*** (0.1217)
Vhat:	3.363*** (0.1218)
Fixed Effects:	
NPI:	Yes
Year:	Yes
S.E. Clustered:	By NPI
Observations:	2,553,058
R2:	0.90926
Within R2:	0.01293

7 Wald Statistic

When using the Anderson-Rubin Wald statistic, I find that the results are same as the traditional t-test. For both, I will reject the null hypothesis.

Table 5: Anderson-Ruben Wald Staistic Test

Dependent Variable:	log(Total Claims)
Practice Revenue Change:	-5e-5*** (1.67e-6)
Fixed Effects:	
NPI:	Yes
Year:	Yes
S.E. Clustered:	By NPI
Observations:	2,252,137
R2:	0.90836
Within R2:	0.00312

The For the second part, I will not need to adjust because my F statistic is above 100. The results are the same then.

8 BH Re-Centering

Table 6: Borusyak and Hull

Dependent Variable:	log(Total Claims)
Intercept:	-3.658*** (0.1984)
Fixed Effects:	
NPI:	Yes
Year:	Yes
S.E. Clustered:	By NPI
Observations:	2,252,137
R2:	0.81485
Within R2:	-1.0141

9 Discussion

When first estimating the effect of integration on the log of claims, the OLS estimate finds a small, significant, negative result. However, the treatment bounds calculated based on varying values of ρ and R^2_{max} show we meed to use an instrumental variable.
Using the 2010 fee schedule update as the instrument, the 2SLS, Durbin-Wu-Hausman test, and Borusyak and Hull all find similar results. There is a large negative relationship between integration status and log of total claims, using the update as an instrument, meaning those who are integrated with the hospital do not claim

10 Reflection

For this assignment, I learned how important separating your code files are. When I went to rerun the code, it recreated the data files, which was completely unnecessary. I need to better organize my files so I do not run into this issue again.

I tried to make my tables more attractive in Latex, but I am still struggling to make effective graphs in R. More practice is required to learn how to edit graphs effectively.

The main takeaway from this assignment is the importance of workflow. Not only do I need to understand how my directory can impact my flow, I also need to learn how to effectively include multiple code files in my workflow.