Module 2 Empirical Exercise

Jingxuan Zhao

In Module 2 Empirical Exercise, I aimed to examine how physician vertical integration status can affect total physician claims. Data on physician claims were obtained from Medicare Utilization and Payment Data. Physician characteristics were based on Medicare Data on Provider Practice and Specialty. Information used to construct the instrumental variable (IV) was from Physician Fee Schedule 2010 Update. Physician with at least 75% of claims billed in an outpatient setting was defined as integrated with a hospital. A log transformation was applied to total physician claims in all the models.

First, I modeled the association of physician integration status with physician claims, controlling for the fixed effects of physician and time and the time-varied physician age. I parameter estimate for physician integration was -0.063 (SE: 0.013) (Table 2). I checked the endogeneity with the method by Oster and found that endogeneity could potentially be an issue for this analysis. Therefore, I constructed the change in Medicare payments achievable for an integrated versus non-integrated physician practice due to the 2010 update to the physician fee schedule and used it as an IV for the analysis. The parameter estimate for physician integration from the Two-Stage Least Squares regression was -3.97 (SE: 0.30) (Table 4), which was quite different from estimates in Table 2. When using the approach from Borusyak and Hull (2021) working paper, the estimate changed to -3.00. In summary, I found that physicians who were integrated to a hospital had lower claims.

Please refer to the ‘readme’ file and R codes for more information on the codes and analysis

Question 1

Table 1. Summary statistics for total physician-level Medicare spending, claims, and patients

	Mean	Standard error	Max	Min
Number of unique patients	865	1195	41480	11
Total claims	1393	2010	132326	11
Total Medicare spending ($)	121360	179616	22439438	29

Question 2

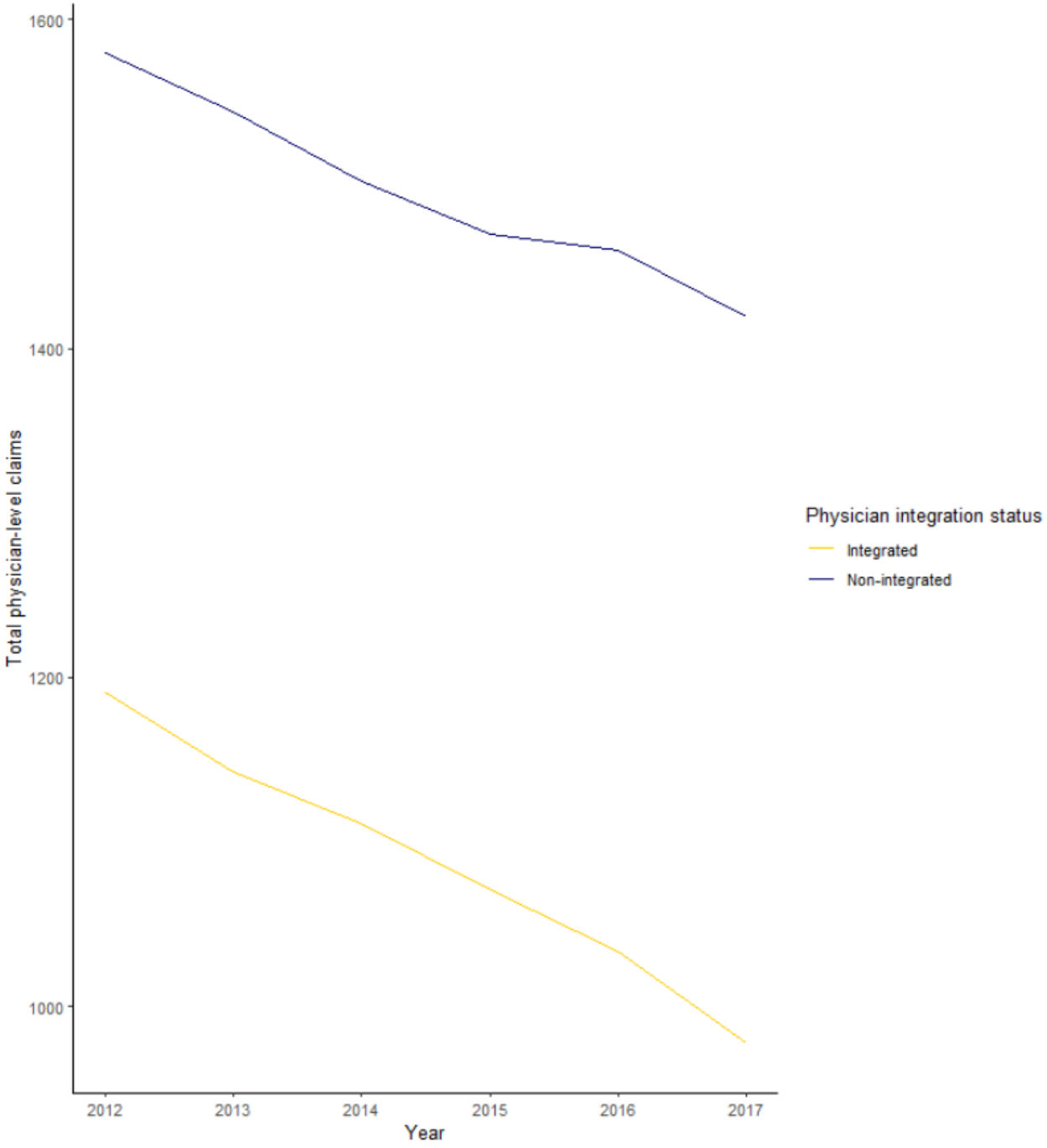

Figure 1. Mean of total physician-level claims for integrated versus non-integrated physicians over time.

Physician with Int≥0.0.75 was classifies as integrated.

From Figure 1, I found that integrated physicians had lower total claims.

Question 3

Table 2. Model summary for physician integration and total Medicare claims

	Model output
Physician integration	-0.063 (0.013)
Age	0.859 (0.508)
Number of observations	1286235
R^2	0.894

For this question, I adjusted for physician age as time-varied variables. The beta estimates for physician integration variable is -0.063 and standard error is 0.013 (p-value<0.05).

Question 4

Table 3. Endogeneity check with methods by Oster (2019)

	R^2_{max}					
ρ	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
0.0	-0.360	-0.360	-0.360	-0.360	-0.360	-0.360
0.5	0.834	1.029	1.213	1.387	1.553	1.713
1.0	2.811	3.477	4.143	4.809	5.475	6.142
1.5	-2.182	-2.182	-2.182	-2.182	-2.182	-2.182
2.0	-2.182	-2.182	-2.182	-2.182	-2.182	-2.182

From Table 3, endogeneity could potentially be an issue for this analysis.

Question 5

Table 4. Model estimates for IV approach, step 1, and the reduced form

	Model output		
	IV model Medicare claims	**Step 1** Physician integration	**Reduced form** Medicare claims
Physician integration	-3.97 (0.30)***		
Change in Medicare payments achievable		$1.50e^{-5}$ ($3.19e^{-7}$)***	$5.98e^{-5}$ ($9.36e^{-7}$)
Age	0.83 (0.003)***	-0.05 (0.22)	1.02 (0.67)
Number of observations	1286235	1286235	1286235
R^2	0.704	0.855	0.894

*** indicates p<0.001

In this IV approach, the parameter estimates for physician integration variable is -3.97 (=$5.98e^{-5}/1.50e^{-5}$) and standard error is 0.30 (p-value<0.05). This is quite different from previous estimate in Table 2.

Question 6

Table 5. Model summary for assessing the need for IV by Durbin-Wu-Hausman test

	Model output
Physician integration	-3.97 (0.10)***
Age	0.83 (0.51)***
Residual	3.92 (0.10)
Number of observations	1286235
R^2	0.894

*** indicates p<0.001

The model summary for $INT_{it} = \lambda \Delta P_{it} + \beta x_{it} + \gamma_i + \gamma_t + \varepsilon_{it}$, is presented in Table 4, step 1.

The parameter estimates for residual is -3.92 and standard error is 0.10 (p-value>0.05), which means there is not enough evidence for the need of IV approach.

I am quite surprised at this finding and thought I may make some mistakes but could not figure it out.

Question 7

Table 6. Results from Anderson-Rubin Wald statistic

	Model output
F statistic	11110.16
P-value	<0.001
95% confidence interval	-1.66 to -1.60

The F-statistic based on 2SLS is 181.1 with P-value <0.001.

The F statistic is 11110.16 with P-value of <0.001. Both this test and t-test following 2SLS estimation of Equation (2) show no evidence of weak IV in this analysis.

Question 8

Table 7. Model estimates for IV approach, step 1, and the reduced form

	Model output		
	IV model Medicare claims	**Step 1** Physician integration	**Reduced form** Medicare claims
Physician integration	-3.00 (0.30)***		
Re-centered change in Medicare payments achievable		$1.32e^{-5}$ ($3.01e^{-7}$)***	$5.29e^{-5}$ ($8.80e^{-7}$)
Age	0.83 (0.003)***	-0.04 (0.23)	1.01 (0.67)
Number of observations	1286235	1286235	1286235
R^2	0.701	0.855	0.894

*** indicates $p<0.001$

Question 9

The findings that physicians who were integrated to a hospital had lower claims were the same as what I expected. At first, I thought once integrated, physicians will get a fixed salary and they will have less motivation to file more claims. After did some literature search, I found this was more complicated than what I thought.

The conclusions were similar from all estimators but the magnitude of the effects was quite different comparing some of them. For example, the estimate from Question 3 is -0.063 before using IV approach. The estimate was -3.97 in Question 5 when using IV in 2SLS. When using the approach from Borusyak and Hull (2021) working paper, the estimate changed to -3.00.

Question 10

One big challenge of this assignment is the large data set. For a real project like this, I will try to find some server resource for data input and cleaning. For question 3, I included state as a time-varied variable at first but I realized that it took a very long time to run the model and also further complicated the rest questions, so I removed it from the model. But I think more covariates will be needed to consider for a real project. I also spent a while for question 4 as physician integration was first coded as a categorical variable, which is not allowed in the o_beta function.

The findings that physicians who were integrated to a hospital had lower claims were the same as what I expected. At first, I thought once integrated, physicians will get a fixed salary and they will have less motivation to file more claims. After did some literature search, I found this was more complicated than what I thought.

In general, I think this assignment helps me better understand the IV approach through applying it to a real analysis.