Final Report for Empirical Exercise 2

Ka Yan CHENG

2022-10-14

Contents

1 Introduction

In this study, we use the Medicare payment shock introduced in 2010 (enhancing payments to physicians for services provided in outpatient facilities compared to those in their offices) as instrument to study the effects of a physician's affiliation with a hospital on physician practice patterns. 3 different datasets are incorporated to implement this study, which are:

1. **Medicare Data on Provider Practice and Specialty -** Details
2. **Medicare Utilization and Payment Data -** Details
3. **Physician Fee Schedule 2010 Update -** Details

Notice that a major limitation of our study is the lack of utilization and payment data before 2012. Our study thus focuses on years from 2012 to 2017. However, even though the price shock was first initiated from 2010, it was not finalized until 2013. Therefore, our study can still take advantage of the gradual variations from 2012 to conduct treatment effect studies as below.

Organization and access to the data are based on practices provided by Prof. Ian McCarthy.

Here is the Github Repo containing all necessary files to reproduce results in this report.

2 Summary Statistics

Provide and discuss a table of simple summary statistics showing the mean, standard deviation, min, and max of total physician-level Medicare spending, claims, and patients.

year	Mean	SD	Min	Max
2012	144503.68	280251.37	3.67	26288557.77
2013	142977.34	281594.79	5.94	20312468.75
2014	147810.90	289625.53	18.73	18764327.50
2015	148792.55	292777.48	0.93	10566763.86
2016	145125.68	299448.23	8.66	18345215.86
2017	149014.96	315564.58	19.16	12834895.39

Table 1: Summary statistics of physician-level Medicare spending

year	Mean	SD	Min	Max
2012	2806.67	11815.09	11.00	2931819.00
2013	2726.88	10591.94	11.00	1122440.00
2014	2960.17	13110.92	4.00	888434.00
2015	3091.67	14296.32	4.40	1314271.00
2016	3022.11	14286.62	5.20	1282350.00
2017	3077.19	14623.43	11.00	1270132.00

Table 2: Summary statistics of physician-level number of claims

year	Mean	SD	Min	Max
2012	1101.64	2189.23	11.00	724713.00
2013	1080.31	1960.66	11.00	721303.00
2014	1079.03	1741.90	11.00	279402.00
2015	1083.42	2205.88	11.00	605766.00
2016	1062.96	1976.15	11.00	635666.00
2017	1051.38	2331.87	11.00	656227.00

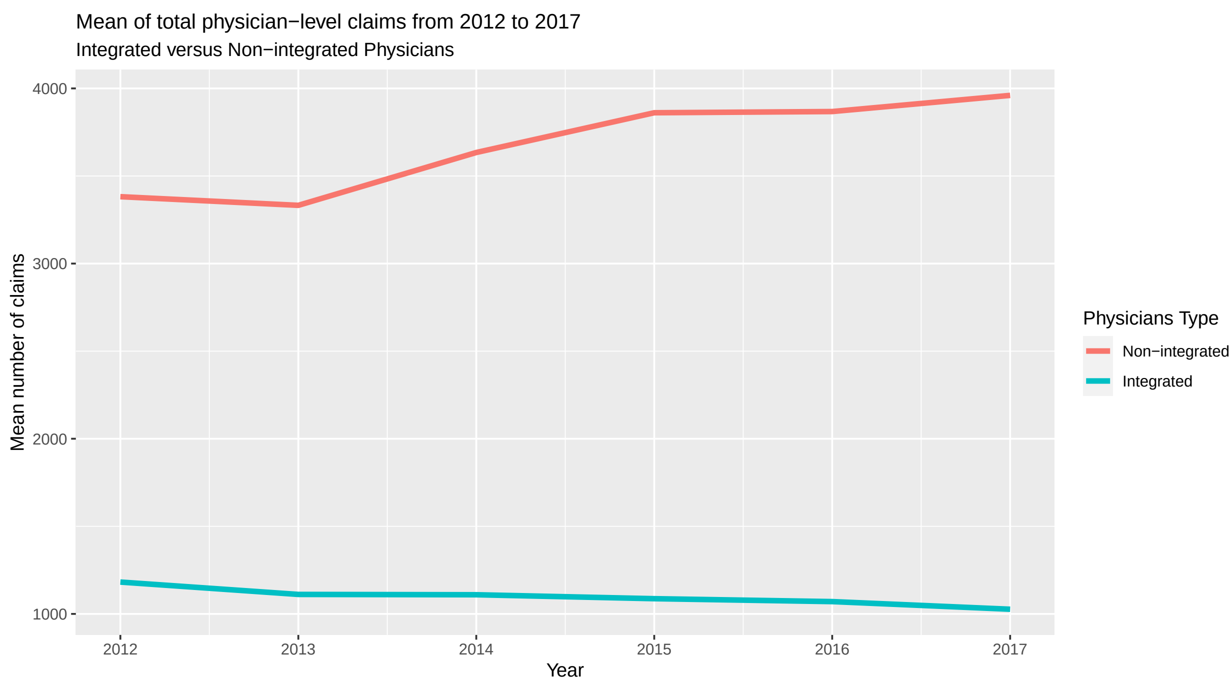
Table 3: Summary statistics of physician-level number of patients

3 Trend of average physician-level claims

Form a proxy for integration using the ratio:

$$INT_{it} = \mathbf{1}\left(\frac{HOPD_{it}}{HOPD_{it} + OFFICE_{it} + ASC_{it}} \geq 0.75\right), \tag{1}$$

where $HOPD_{it}$ reflects the total number of claims in which physician i bills in a hospital outpatient setting, $OFFICE_{it}$ is the total number of claims billed to an office setting, and ASC_{it} is the total number of claims billed to an ambulatory surgery center. As reflected in Equation (1), you can assume that any physician with at least 75% of claims billed in an outpatient setting is integrated with a hospital. Using this 75% threshold, plot the mean of total physician-level claims for integrated versus non-integrated physicians over time.



4 Basic TWFE Model

Estimate the relationship between integration on (log) total physician claims using OLS, with the following specification:

$$y_{it} = \delta INT_{it} + \beta x_{it} + \gamma_i + \gamma_t + \varepsilon_{it}, \tag{2}$$

where INT_{it} is defined in Equation (1), x_{it} captures time-varying physician characteristics, and γ_i and γ_t denote physician and time fixed effects. Please focus on physician's that weren't yet integrated as of 2012, that way we have some pre-integration data for everyone. Impose this restriction for the remaining questions. Feel free to experiment with different covariates in x_{it} or simply omit that term and only include the fixed effects.

	Basic TWFE Model
Integration	−0.337***
	(0.007)
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

5 Oster bound

How much should we be "worried" about endogeneity here? Extending the work of @altonji2005, @oster2019 derives the expression

$$\delta^* \approx \hat{\delta}_{D,x_1} - \rho \times \left[\hat{\delta}_D - \hat{\delta}_{D,x_1}\right] \times \frac{R^2_{max} - R^2_{D,x_1}}{R^2_{D,x_1} - R^2_D} \xrightarrow{p} \delta, \tag{3}$$

where x_1 captures our observable covariates (or fixed effects in our case); δ denotes the treatment effect of interest; $\hat{\delta}_{D,x_1}$ denotes the coefficient on D from a regression of y on D and x_1; R^2_{D,x_1} denotes the R^2 from that regression; $\hat{\delta}_D$ denotes the coefficient on D from a regression of y on D only; R^2_D reflects the R^2 from that regression; R^2_{max} denotes an unobserved "maximum" R^2 from a regression of y on D, observed covariates x_1, and some unobserved covariates x_2; and ρ denotes the degree of selection on observed variables relative to unobserved variables. One approach that Oster suggests is to consider a range of R^2_{max} and ρ to bound the estimated treatment effect, where the bounds are given by $\left[\hat{\delta}_{D,x_1}, \delta^*(R^2_{max}, \rho)\right]$. Construct these bounds based on all combinations of $\rho \in (0, .5, 1, 1.5, 2)$ and $R^2_{max} \in (0.5, 0.6, 0.7, 0.8, 0.9, 1)$ and present your results in a table. What do your results say about the extent to which selection on observables could be problematic here?

	0.5	0.6	0.7	0.8	0.9	1
0.5	[-0.34,1.14]	[-0.34,1.43]	[-0.34,1.73]	[-0.34,2.02]	[-0.34,2.32]	[-0.34,2.61]
1	[-0.34,-0.78]	[-0.34,-0.86]	[-0.34,-0.95]	[-0.34,-1.04]	[-0.34,-1.13]	[-0.34,-1.21]
1.5	[-0.34,-2.69]	[-0.34,-3.16]	[-0.34,-3.63]	[-0.34,-4.1]	[-0.34,-4.57]	[-0.34,-5.04]
2	[-0.34,-4.6]	[-0.34,-5.46]	[-0.34,-6.31]	[-0.34,-7.16]	[-0.34,-8.02]	[-0.34,-8.87]

Table 4: Oster bound of the estimated treatment effect

6 IV version of the TWFE model

Construct the change in Medicare payments achievable for an integrated versus non-integrated physician practice due to the 2010 update to the physician fee schedule, ΔP_{it}. Use this as an instrument for INT_{it} in a 2SLS estimator following the same specification as in Equation (2). Present your results along with those of your "first stage" and "reduced form".

	First stage	Reduced form	TSLS
Price change	0.000***	0.000***	
	(0.000)	(0.000)	
Integration			−13.065***
			(0.726)

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

7 Durbin-Wu-Hausman test

Assess the "need" for IV by implementing a Durbin-Wu-Hausman test with an augmented regression. Do this by first estimating the regression, $INT_{it} = \lambda \Delta P_{it} + \beta x_{it} + \gamma_i + \gamma_t + \varepsilon_{it}$, take the residual $\hat{\nu} = INT_{it} - \hat{INT}_{it}$, and run the regression

$$y_{it} = \delta INT_{it} + \beta x_{it} + \gamma_i + \gamma_t + \kappa \hat{\nu} + \varepsilon_{it}.$$

Discuss your results for $\hat{\kappa}$.

	Durbin-Wu-Hausman test
Integration	-13.065***
	(0.728)
Residual	-0.311***
	(0.008)

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

8 Weak instrument tests

Now let's pay attention to potential issues of weak instruments. As we discussed in class, one issue with weak instruments is that our typical critical values (say, 1.96 for a 95% confidence interval) from the equation of interest (sometimes called the structural equation) are too low in the presence of a weak first-stage. These issues are presented very clearly and more formally in the Andrews, Stock, and Sun (2019) survey article. For this question, you will consider two forms of inference in the presence of weak instruments. Hint: Check out the `ivmodel` package in R or the `ivreg2` command in Stata for help getting the AR Wald statistic.

- Present the results of a test of the null, $H_0 : \delta = 0$, using the Anderson-Rubin Wald statistic. Do your conclusions from this test differ from a traditional t-test following 2SLS estimation of Equation (2)?

lower	upper
-30.26	-8.17

Table 5: Anderson-Rubin confidence interval

- Going back to your 2SLS results… inflate your 2SLS standard errors to form the tF adjusted standard error, following Table 3 in @lee2021. Repeat the test of the null, $H_0 : \delta = 0$, using standard critical values and the tF adjusted standard error.

We shall observe that the F-statistics in the first stage is big (102.67), so adjustment is not effective. Below is the confidence interval for the coefficient in the 2SLS model.

lower	higher
-13.74	-12.39

Table 6: 2SLS confidence interval

9 Borusyak and Hull (2021) re-centering approach

Following the Borusyak and Hull (2021) working paper (BH), we can consider our instrument as a function of some exogenous policy shocks and some possibly endogenous physician characteristics, $\Delta P_{it} = f\left(g_{pt}; z_{ipt}\right)$, where g_{pt} captures overall payment shocks for procedure p at time t, and z_{ip} denotes a physician's quantity of different procedures at baseline. We can implement the BH re-centering approach as follows: - Consider hypothetical price changes over a set of possible counterfactuals by assuming that the counterfactuals consist of different allocations of the observed relative price changes. For example, take the vector of all relative price changes, reallocate this vector randomly, and assign new hypothetical relative price changes. Do this 100 times. This isn't "all" possible counterfactuals by any means, but it will be fine for our purposes. - Construct the expected revenue change over all possible realizations from previously, $\mu_{it} = E[\Delta P_{it}] = \sum_{s=1}^{100} \sum_p g_{pt}^s z_{ip}$. - Re-estimate Equation (2) by 2SLS when instrumenting for INT_{it} with $\tilde{\Delta} P_{it} = \Delta P_{it} - \mu_{it}$. Intuitively, this re-centering should isolate variation in the instrument that is only due to the policy and remove variation in our instrument that is due to physician practice styles (the latter of which is not a great instrument).

	BH re-centering approach
Integration	−12.852***
	(0.715)

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

10 Discussion

Discuss your findings and compare estimates from different estimators.

11 Reflection

Reflect on this assignment. What did you find most challenging? What did you find most surprising?