

# Econ771 - Empirical Exercise 2

Nixon Torres Candiales

October 15, 2022

## Overview

In this assignment, we're going to work through some applied issues related to instrumental variables. For a long time, IV (or 2SLS) was a very common identification strategy for applied empirical micro, but it fell out of favor as people became more aware of the assumptions underlying the estimator and better understood what IV actually estimates (not the ATE in most cases). People also started to find other strategies that were more compelling in some applications (and of course with some other assumptions). In this assignment, we're going to study the effects of a physician's affiliation with a hospital on physician practice patterns, and we'll instrument for physician affiliation using some specific Medicare payment shocks.

Please "submit" your answers as a GitHub repository link. In this repo, please include a final document with your main answers and analyses in a PDF. Be sure to include in your repository all of your supporting code files. Practice writing good code and showing me only what I would need to recreate your results.

## Resources and data

The data for this assignment comes from three sources:

1. MD-PPAS; The Medicare Data on Provider Practice and Specialty includes data on physician specialties, practice IDs, demographics, and place of service. Be sure to follow the link and read the data documentation. We'll use these data to construct a measure

of physician integration.

2. Medicare Utilization and Payment Data: These files provide data on the quantities and Medicare spending of each physician and service. We'll use these data to capture total physician-level billing activity, and we'll use the service-level data to measure the revenue effects from our plausibly exogenous policy shock. These data are only available beginning in 2012. These files are large but otherwise relatively clean and easy to use, so there's no separate repo for these data. Note that we will only work with data for MDs, so you can drop a lot of observations with that restriction.
3. Physician Fee Schedule 2010 Update: Our instrument mainly consists of a shock to physician payments introduced in 2010. The shock further increased payments for services in an outpatient facility compared to services billed in a physician's office. The GitHub repo (linked above) provides code to recreate a dataset with service-specific price shocks introduced by the 2010 fee schedule update. To save us some time, I've posed the final dataset from that repo into our class data folder.

## Questions

1. 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. Use the Medicare utilization and payment data to calculate total spending, claims, and patients at the physician level. The patient counts will include some overlap since the data are by service, but that's OK for our purposes.

	Mean	Std.Dev.	Min	Max
Spending	137743.61	280251.91	0.93	26288557.77
Claims	2699.64	13038.64	4.00	5750425.00
Patients	1013.11	1906.22	11.00	724713.00

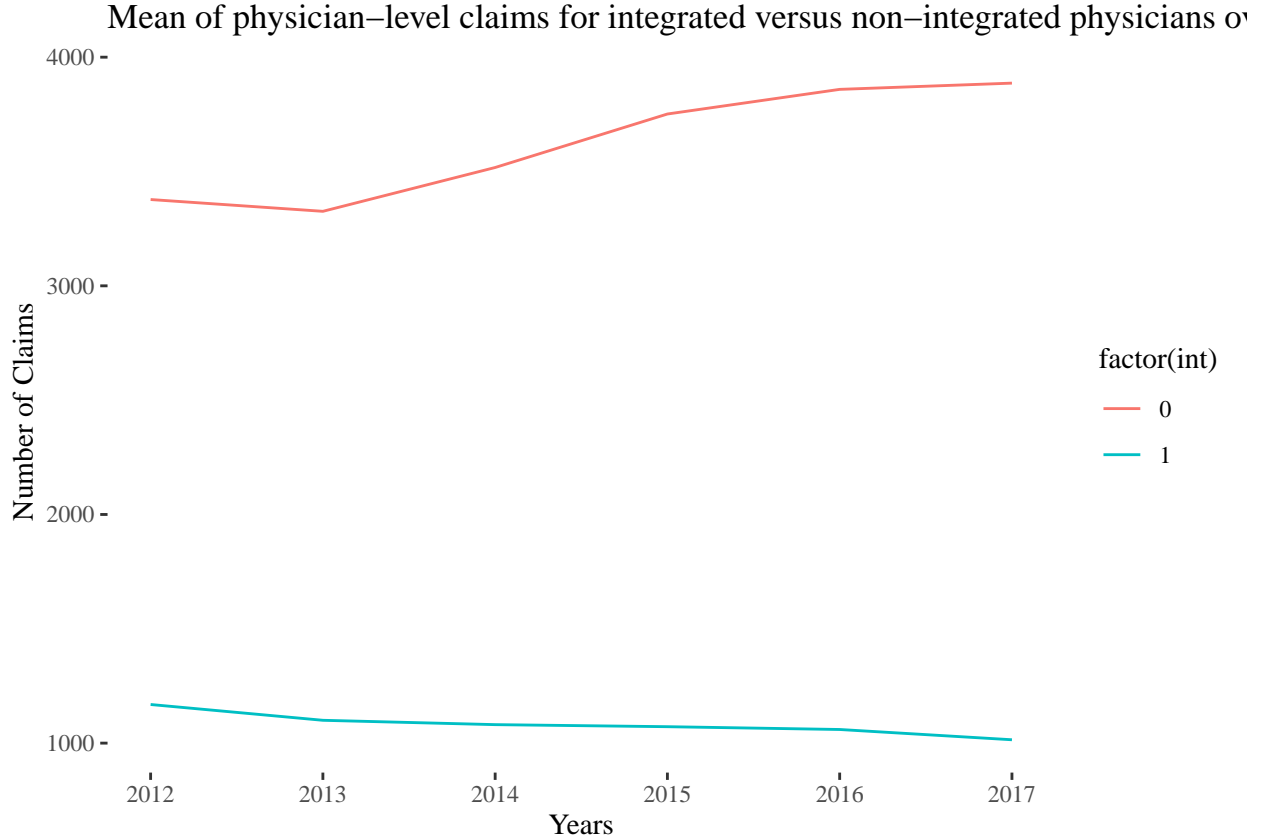
Table 1: Summary statistics

Table 1 shows the summary statistics for the total spending, number of claims and total of patients over the years of interest from 2012 to 2017.

2. Form a proxy for integration using the ratio:

$$INT_{it} = \mathbf{1} \left( \frac{HOPD_{it}}{HOPD_{it} + OFFICE_{it} + ASC_{it}} \geq 0.75 \right), (\#eq : int) \quad (1)$$

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 @ref(eq:int), 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.



We see that after the price shock implementation the total number of claims increased for unintegrated physicians, whereas for the integrated physicians we do not observe a change in the trend.

3. Estimate the relationship between integration on total physician claims using OLS,

with the following specification:

$$y_{it} = \delta INT_{it} + \beta x_{it} + \gamma_i + \gamma_t + \varepsilon_{it}, (\#eq : ols) \quad (2)$$

where  $INT_{it}$  is defined in Equation @ref(eq:int),  $x_{it}$  captures time-varying physician characteristics, and  $\gamma_i$  and  $\gamma_t$  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.

	log $y$ (1)
average submitted charge	$1.04 \times 10^{-5***}$ ( $7.89 \times 10^{-7}$ )
average medicare payment	$0.0002***$ ( $7.6 \times 10^{-6}$ )
$INT$	$-0.2624***$ (0.0047)
Observations	2,565,310
R <sup>2</sup>	0.91054
Within R <sup>2</sup>	0.12193
npi fixed effects	✓
Year fixed effects	✓

Table 2: OLS estimation

The estimation results indicate that integrated physicians submit 0.26% less claims than the unintegrated physicians.

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

$$\delta^* \approx \hat{\delta}_{D,x_1} - \rho \times [\hat{\delta}_D - \hat{\delta}_{D,x_1}] \times \frac{R_{max}^2 - R_{D,x_1}^2}{R_{D,x_1}^2 - R_D^2} \xrightarrow{p} \delta, (\#eq : oster) \quad (3)$$

where  $x_1$  captures our observable covariates (or fixed effects in our case);  $\delta$  denotes

the treatment effect of interest;  $\hat{\delta}_{D,x_1}$  denotes the coefficient on  $D$  from a regression of  $y$  on  $D$  and  $x_1$ ;  $R_{D,x_1}^2$  denotes the  $R^2$  from that regression;  $\hat{\delta}_D$  denotes the coefficient on  $D$  from a regression of  $y$  on  $D$  only;  $R_D^2$  reflects the  $R^2$  from that regression;  $R_{max}^2$  denotes an unobserved “maximum”  $R^2$  from a regression of  $y$  on  $D$ , observed covariates  $x_1$ , and some unobserved covariates  $x_2$ ; and  $\rho$  denotes the degree of selection on observed variables relative to unobserved variables. One approach that Oster suggests is to consider a range of  $R_{max}^2$  and  $\rho$  to bound the estimated treatment effect, where the bounds are given by  $[\hat{\delta}_{D,x_1}, \delta^*(R_{max}^2, \rho)]$ . Construct these bounds based on all combinations of  $\rho \in (0, .5, 1, 1.5, 2)$  and  $R_{max}^2 \in (0.5, 0.6, 0.7, 0.8, 0.9, 1)$  and present your results in a table. What do your results say about the extent to which selection on observables could be problematic here? Hint: you can also look into `psacalc` in **Stata** or `robomit` in **R** for implementation of @oster2019 in **Stata** or **R**, respectively.

	$R_{max}^2 = 0.9$	$R_{max}^2 = 1$
$\rho = 0$	[-0.36,-0.36]	[-0.36,-0.36]
$\rho = 0.5$	[-0.36,-0.36]	[-0.36,-0.33]
$\rho = 1$	[-0.36,-0.36]	[-0.36,-0.31]
$\rho = 1.5$	[-0.36,-0.36]	[-0.36,-0.29]
$\rho = 2$	[-0.36,-0.36]	[-0.36,-0.26]

Table 3: Altonji, Elder, and Taber (2005)

used (Mb) gc trigger (Mb) limit (Mb) max used (Mb) Ncells 4019019 214.7 12889335 688.4  
NA 9838106 525.5 Vcells 48186612 367.7 145748861 1112.0 102400 145598160 1110.9

From the regression presented in table 1 we see that the  $R_{D,x_1}^2$  was 0.91. Thus, we omit  $R_{max}^2 \leq 0.9$ . In all the simulations the bounds remain negative. Thus

5. 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,  $\Delta P_{it}$ . Use this as an instrument for  $INT_{it}$  in a 2SLS estimator following the same specification as in Equation @ref(eq:ols). Present your results along with those of your “first stage” and “reduced form”.

Yes, the idea of summing a ratio is a bit odd. But it’s easier to think of the instrument as the product of baseline (pre-shock) practice size and the average relative revenue change due

to the price shock. In that context, the sum of the ratio is really just an interaction term that incorporates information on the price shock magnitude and baseline practice size. Each of these things alone are poor instruments, but together for the practice it reflects a “better” instrument.

	$INT$ (1)	$\log y$ (2)	(3)
practice revenue change	$1.33 \times 10^{-5***}$ ( $6.76 \times 10^{-7}$ )	$-4.29 \times 10^{-5***}$ ( $1.57 \times 10^{-6}$ )	
average submitted charge		$1.02 \times 10^{-5***}$ ( $9.29 \times 10^{-7}$ )	$1.02 \times 10^{-5***}$ ( $9.29 \times 10^{-7}$ )
average medicare payment		$0.0003***$ ( $8.83 \times 10^{-6}$ )	$0.0003***$ ( $8.83 \times 10^{-6}$ )
$\hat{INT}$			$-3.233***$ (0.1184)
Observations	2,176,278	2,176,278	2,176,278
R <sup>2</sup>	0.89170	0.92032	0.92032
Within R <sup>2</sup>	0.00235	0.12169	0.12169
npi fixed effects	✓	✓	✓
Year fixed effects	✓	✓	✓

Table 4: Two Stage Least Squares estimation

Endogeneity can be a threat for identification if the number of claims affect the decision to merge. Thus, table 4 presents the results of the two stage least squares. We observe how the estimates decrease, indicating that the number of claims for unintegrated physicians further decline given the price shock.

6. 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  $\hat{\kappa}$ .

	log $y$ (1)
<i>INT</i>	-3.444*** (0.1253)
average submitted charge	$1.3 \times 10^{-5}$ *** ( $9.31 \times 10^{-7}$ )
average medicare payment	0.0002*** ( $9.66 \times 10^{-6}$ )
$\hat{v}$	3.204*** (0.1251)
Observations	2,176,278
R <sup>2</sup>	0.92079
Within R <sup>2</sup>	0.12692
npi fixed effects	✓
Year fixed effects	✓

Table 5: Durbin-Wu-Hausman Test

From the Durbin-Wu-Hausman test with an augmented regression we see how  $\hat{v}$ , the residuals, is significant which implies that there was a systematic component in the error that has not been taken care of and this is also driving the treatment effect. As such, this results might be bias.

7. 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:

- 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 @ref(eq:ols)?
- Going back to your 2SLS results... inflate your 2SLS standard errors to form the  $tF$

adjusted standard error, following Table 3 in Lee et al. (2021). Repeat the test of the null,  $H_0 : \delta = 0$ , using standard critical values and the  $tF$  adjusted standard error.

	lower	upper
Anderson-Rubin	-4.11	-3.84
Lee	-4.07	-3.87

Table 6: Confidence intervals

The first stage F statistic is 4013. Then, there is no evidence of weak instruments, as presented in the table.

8. 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,  $\Delta P_{it} = f(g_{pt}; z_{ipt})$ , 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 @ref(eq:ols) 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).

Following Borusyak and Hull (2021) working paper, we recenter the instrument variable by the expected value of the pseudo instruments simulations. After re-centering the instrument the treatment effect shrinks but we still have a larger negative effect than in table 2.

9. Discuss your findings and compare estimates from different estimators.



	$\log y$ (1)
$I\hat{N}T$	-0.8217*** (0.1355)
Observations	2,176,278
R <sup>2</sup>	0.90929
Within R <sup>2</sup>	0.00011
npi fixed effects	✓
Year fixed effects	✓

Table 7: Borusyak and Hull (2021) estimation

After all of the analysis and sensitivity tests we can see how the integration to the hospital had a negative effect on the number of claims physician made.

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

These empirical assignments are a good opportunity to polish the workflow needed to advance our own research. I have enjoyed them all so far. I found working with Rmarkdown for the PDF the most challenging part so far. In this particular case, as the objects in memory increased I was constrained by the capacities of my device, and as such, I am considering on writing the results in a text file and loading it later, as is the case when I use latex for the write-up.

But is a good practice to polish my workflow and improve the reproducibility of my research and the automation of the tables.