Martha Wetzel

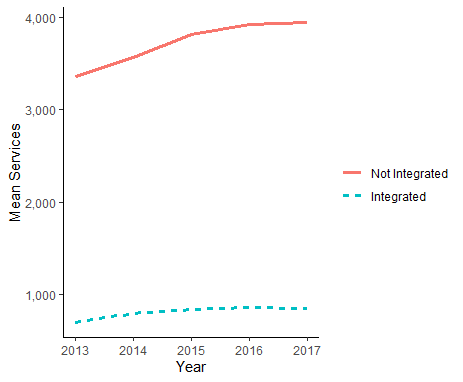
ECON 771

Assignment 2

1. The table below describes physician-level Medicare spending, services, and patients over the years 2012-2017. The median number of Medicare beneficiaries seen by a physician in a year was 515. Amount allowed, services, and beneficiaries were all heavily skewed.

| **Variable** | **Mean** | **Standard Deviation** | **Minimum** | | **50th Percentile (Median)** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| Total Amt Allowed | 136,570 | 278,311 | 1 | | 67,153 | 26,288,558 |
| Number of Services | 2,676 | 12,863 | 4 | | 786 | 5,750,425 |
| Number of Medicare Beneficiaries | 1,014 | 1,933 | 11 | | 515 | 724,700 |
| Summary statistics based on 3,012,951 physician-years. | | | |

1. The figure below shows the mean number of services per physician per year, grouped by integration status. The integrated physicians have lower service counts than the non-integrated physicians.



1. The table below summarizes the results from an OLS regression of logged services on integration status. This shows that integration is associated with a 33% decrease in services provided.

| **Variable** | **OLS** |
| --- | --- |
| Integration | 0.67 |
|  | [0.66, 0.68] |
| Age | 0.91 |
|  | [0.73, 1.15] |
| Num.Obs. | 2080213 |
| Std.Errors | by: npi |
| FE: npi | X |
| FE: Year | X |
| *Note: Exponentiated coefficients presented with 95% CIs.* | |

1. The bounds on the raw estimate of -0.4 are shown below. Either I did this wrong or we should be extremely concerned. Conceptually, I’m not sure this test makes the most sense when my only exogenous variable is age, which was not significantly related to services rendered in the previous regression. It seems like you would need to have some actually predictive exogenous variables in order to claim that the selection on observables is informative about selection on unobservables.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **R-Squared Values** | **Range of Deltas** | | | |
| **0.5** | **1** | **1.5** | **2** |
| 0.5 | 1.59 | 8.52 | -2.36 | -2.36 |
| 0.6 | 1.95 | 10.57 | -2.36 | -2.36 |
| 0.7 | 2.28 | 12.62 | -2.36 | -2.36 |
| 0.8 | 2.58 | 14.68 | -2.36 | -2.36 |
| 0.9 | 2.87 | 16.73 | -2.36 | -2.36 |
| 1 | 3.15 | 18.78 | -2.36 | -2.36 |

1. The following table displays the two stage least squares (2SLS) estimates, along with the first stage and reduced form estimates. The 2SLS estimate suggests that integration is associated with an increase in services provided, which is the opposite of the finding from the regression of services on integration without the instrument.

| **Variable** | **2SLS: Services on Integration** | **Stage 1: Integration on Expected Revenue** | **Reduced Form: Services on Expected Revenue** |
| --- | --- | --- | --- |
| Integration | 6.73 [4.79, 9.47] |  |  |
| Age | 0.8 [0.54, 1.17] | 0.98753 [0.9482, 1.02848] | 0.7779 [0.5683, 1.0649] |
| Expected Revenue Change |  | 1.00004 [1.00004, 1.00004] | 1.0001 [1.0001, 1.0001] |
| Num.Obs. | 1126501 | 1126501 | 1126501 |
| Std.Errors | by: npi | by: npi | by: npi |
| FE: npi | X | X | X |
| FE: Year | X | X | X |
| *Note: Exponentiated coefficients presented with 95% CIs.* | | | |

1. The estimate for is 1.00 (SD: 0.00003). This indicates that the OLS estimates are inconsistent.
2. I couldn’t figure out how to get the Anderson-Rubin Wald statistic for the 2SLS with fixed effects. When I ran it on the model without the fixed effects, I got a test statistic of 4343 (p-value < 2.22e-16), so it didn’t change that estimate, but also, that estimate was way different due to the omission of the fixed effects. Going back to the 2SLS results, the first-stage f-statistic was already 1,454. This is such a high f-statistic that no adjustment was needed.
3. …I didn’t finish this one
4. My estimates are completely different depending on the method used. The finding from 2SLS that integration increases services was surprising, since the integrated physicians raw service counts were much lower than the non-integrated physicians. I’m not familiar enough with IV to know what to expect from the first stage vs reduced form vs 2SLS, and so it is hard for me to judge if these results make sense or not.
5. I really struggled to figure out how to do IV with fixed effects in a way that would produce some of the requested statistics without breaking R. I think if the data had been smaller, I might have been able to just include dummy variables for NPI and year, but not with data of this size. I found all of the results surprising! I felt like I was doing the programming wrong to be getting these results.