Example 2: dynamic inefficiencies in employment-based health insurance

Paul Schrimpf

UBC
Economics 326

March 24, 2015

Example 2: dynamic inefficiencies in employment-based health insurance

Paul Schrimpf

Plan

- Last week and today: how to use the methods we've studied
- Today: another walk-through research project
- This time, we will focus less on big picture issues (how to choose a topic, etc) and more on details and interpreting results

References

- “Dynamic Inefficiencies in an Employment-Based Health Insurance System: Theory and Evidence” Fang and Gavazza (2011)

Section 1

Introduction

Introduction

- US health care:
  - Private, employment-based insurance (55%) [1]
  - Private direct purchase insurance (9.8%)
  - Retirees over 65: Medicare (14.5%)
  - Low income : Medicaid (15.9%)
  - Uninsured (16.3%)

[1] These numbers come from table 10 of DeNavas-Walt, Proctor, and Smith (2011). Some people have more than one form of insurance, so these percentages need not add to 100.

Introduction

- Basic idea:
  - Health is a form of human capital
  - Firms and workers will want to invest in workers' health
  - Possibility of worker turnover can lead to under investment in health
  - Under investment in health while working $\Rightarrow$ higher health spending in retirement

Theoretic Predictions

- Workers in jobs with lower turnover rates have:
  1. Higher medical expenditures while working
  2. Lower medical expenditures and better health during retirement
- Fang and Gavazza (2011) show that these are the implications of a simple model with labor market frictions (jobs end and part of the higher wage associated with better health is lost) and employer provided health investment

Section 2

Data

Data

- Medical Expenditure Panel Survey (MEPS) household component 1995-2005
  - Rotating panel (follows people for 2 years)
  - Annual medical expenditure, number of doctor visits
  - Job tenure, job industry, income
  - Demographics: age, education, gender, race, marital status, household size, geographic region

Data

Scenario: we have gathered our data. What is the first thing we should do with it?

We should check that it makes sense. Let's look at some tables and figures. What tables and figures should we create?

Section 3

Empirical specification

Empirical specification

- Want to test whether workers in jobs with lower turnover rates have:
  1. Higher medical expenditures while working
  2. Lower medical expenditures and better health during retirement (we will look at this next)
- What equation do we want to estimate?

OLS

$$\log(spend_{it}) = \beta_0 + \beta_1 \log(Tenure_{it}) + \beta_x X_{it} + \\ + \text{region} \times \text{year dummies} + \zeta_i + \epsilon_{it}$$

- What assumption do we need for OLS to be consistent?
- Suppose OLS is consistent, what is the predicted sign of $\hat{\beta}_1^{OLS}$?
- Why is OLS likely not consistent? In what direction do you think OLS will be biased?

OLS - interpretation

How can we interpret the OLS estimates? Does the coefficient of interest have the expected sign? Is it small or large?

OLS - inference

How should we calculate standard errors?

IV

We need an instrument for job tenure. What conditions does an instrument need to satisfy? What could be a good instrument?

IV

We need an instrument for job tenure. What conditions does an instrument need to satisfy? What could be a good instrument?

We will use the number of establishments that closed and jobs that disappeared in a person's region, industry and year as instruments. What is the rationale for this instrument? Are there any reasons to think it is not a valid instrument?

IV

Establishment closures are a valid instrument if:

1. Relevant: "The employer-employee pair forms expectations about plant closures that are correlated with their realizations, so that expected turnover and actual turnover generated by plant closures covary"
2. Exogenous: "Establishment deaths do not directly affect individual medical expenditures—i.e., the exclusion restriction"

2SLS - specification

How should we specify the 2SLS estimator? What is the dependent variable(s)? What is the endogenous regressor of interest? What is the instrument? What controls should we include?

What is the first stage? What is the reduced form? What should we check in these regressions?

Code and data

- Code for main results
- Code for calculating heteroskedasticity robust clustered standard errors
- Data

References

DeNavas-Walt, Carmen, Bernadette D Proctor, and Jessica C Smith. 2011. "Income, poverty, and health insurance coverage in the United States: 2010." *Current Population Reports: Consumer Income* URL http://www.census.gov/prod/2011pubs/p60-239.pdf.

Fang, Hanming and Alessandro Gavazza. 2011. "Dynamic inefficiencies in an employment-based health insurance system: Theory and evidence." *The American Economic Review* 101 (7):3047–3077. URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.7.3047.