PP346 Program Evaluation

Problem Set 1   **Due Tuesday, October 16, by 9 am**

For this assignment, provide a write-up where you answer the questions below, selectively cutting and pasting output where needed. Be concise in your write-up; excess wordiness will be penalized. Also submit a log file that includes commands and results for your entire analysis.

Study motivation

Developing and middle-income countries increasingly provide health coverage to their residents. Mexico has become a leading example. Prior to 2004, roughly half the Mexican population had no health insurance. Health coverage was available through the social security system for salaried employees and their families, but the roughly 50 million Mexicans outside the salaried sector had no insurance.

In 2004, Mexico introduced Seguro Popular. Seguro Popular is a voluntary, non-contributory health insurance program for residents without coverage from the social security system. One of its explicit goals is to provide covered households with financial protection by reducing their incidence of catastrophic out-of-pocket health spending.

The Seguro Popular field experiment

The experiment involved a paired-cluster randomization design, by which 100 largely rural "health clusters" were formed into 50 pairs on the basis of similarities in demographic characteristics and healthcare infrastructure. Health clusters are defined as a healthcare facility and the population living within its catchment area. Within matched pairs of health clusters, one was randomly assigned to treatment, with the other serving as control. Treatment involved publicity campaigns designed to encourage local residents to enroll. Extensive information about the initial design of the experiment has been published by King et al (2009).

Pre-intervention baseline data were collected from participant households in 2005. Post-intervention follow-up data were collected in 2006, roughly 10 months after the intervention began. The 2005 survey included 32,515 households.

Questions:

(1) Did Seguro Popular reduce extreme expenditures on health?

(2) Does it matter how we define "extreme expenditures?"

Data

Data for this assignment are in the file SP_dataset.dta on Canvas.

Analysis:

(1) Inspect the data. Are there truly 100 clusters in 50 pairs? Does each pair have a treatment and a control cluster? What is the modal level of schooling in the sample? The modal marital status? What share of sample households were participating in Oportunidades, Mexico's conditional cash transfer program? What is mean non-health expenditure (data are in Mexican pesos, where 12 pesos = approx. 1 dollar)? Does this make sense?

(2) Which treatment parameters could be identified by these data?

(3) Assess the adequacy of random assignment. You will need to think about how to use Stata's reg command to test for equal means among pre-intervention variables within cluster pairs. How important is it to account for dependence within clusters? What is the bottom line? Was random assignment properly executed?

(4) Construct two variables relating health spending to the household's total income. The first is the health budget share, equal to health spending divided by total spending. The second is the adjusted health budget share, equal to health spending divided by disposable income, where disposable income is defined as total spending less food expenditures.

(5) On average, what share of total expenditures does health spending account for? What are the 75th and 90th percentiles of the budget share distribution? What are the corresponding figures as a share of disposable income rather than total expenditures?

(6) Construct two measures of extreme health spending: (i) a dummy equal to one if the household's health budget share exceeds 20 percent; and (ii) a dummy equal to one if the household spends more than 30 percent of its disposable income on health. How are these measures related?

(7) Estimate the ITT effects of the intervention on extreme health spending. First, just regress the extreme health spending measures on the treatment dummy. Next, add cluster-pair dummies. In both cases, calculate standard errors that account for dependence among households within clusters. What happens to the coefficients when you add the cluster-pair dummies? To their standard errors? Why?

(8) Now adjust for baseline characteristics. Again, do this with and without the cluster-pair dummies in the regression. What happens to the estimated treatment effects? To their standard errors?

(9) Are any of your conclusions sensitive to the way that you measure extreme health expenditures?

(10) Now calculate the Wald estimator. How do you interpret it? What additional assumptions do you need to invoke for these estimates to be valid?