Program Evaluation with Panel Data

Kampala RPC Workshop, Day 1

January 27, 2008

Ms. Speedy Analyst's Problem

Differences in Differences

Propensity Score Matching

Ms. Speedy Analyst's Problem

The Mystery of the Vanishing Benefits:

▶ PROSCOL is a government anti-poverty program. It pays cash transfers to poor families who keep their children in school.

▶ Before deciding whether to expand the program, the government of Labas wants to know whether PROSCOL works or not.

▶ Ms. Speedy analyst has data on program participants as well as some non-participants.

▶ Sadly, perhaps, schooling levels for these two groups are not noticeably different.

Ms. Speedy Analyst's Problem

We're interested in measuring the **treatment effect** of the PROSCOL program, or the causal impact of participation in the program:

$$G = E(S_{1i} - S_{0i}|P_i = 1) \tag{1}$$

In words: for a person who chose to participate (i.e., conditional on $P_i = 1$), the treatment effect is the difference between their schooling level obtained in the program and what they would have attained otherwise.

This is not the same as comparing schooling for those who did and did not participate, which gives

$$
\begin{aligned}
D &= E(S_{1i}|P_i = 1) - E(S_{0i}|P_i = 0) \tag{2} \\
&= G + B \tag{3}
\end{aligned}
$$

This difference D is equal to the treatment effect G plus a bias B which we can derive as

$$B = E(S_{0i}|P_i = 1) - E(S_{0i}|P_i = 0) \tag{4}$$

Ms. Speedy Analyst's Problem

Speedy tries to solve her problem by running the following regression:

$$S_i = a + bP_i + cX_i + \varepsilon_i \qquad (5)$$

where P is a dummy for participation and X is a vector of household and/or village characteristics. b measures the treatment effect. Her estimate of the ith household's schooling is

$$a + bP_i + cX_i + \varepsilon_i \qquad \text{if it participates} \qquad (6)$$
$$a + bP_i + \varepsilon_i \qquad \text{if it does not} \qquad (7)$$

Thus Speedy implicitly assumes ε_i is the same in both cases. This is equivalent to assuming ε is independent of P. More intuitively, it assumes that she has controlled for everything influencing participation that might also affect schooling.

Suppose Ms. Speedy had a baseline survey...

What is panel data?

▶ Panel data combines a cross-section dimension (N households) and a time-series dimension (T periods); that is, we follow a large group of observations over time.

▶ Distinguish panel data from repeated or pooled cross-section data: panels follow the same households (or individuals, districts, firms, etc.) while pooled cross-sections re-sample each period.

Panel data is relatively expensive to collect. Why bother?

▶ More observations $=$ more precise estimates

▶ We can allow for time trends and time-specific effects (annual shocks or seasonality)

▶ More importantly, panel data can solve an endogeneity problem that leads to erroneous conclusions in policy evaluation.

Suppose Ms. Speedy had a baseline survey. . .

Professor Chisqure suggests that Speedy rewrite her regression equation as follows:

$$S_{it} = a + bP_{it} + cX_{it} + \varepsilon_{it} \qquad (8)$$

for $t = A, B$. If we write an explicit equation for this error term

$$\varepsilon_{it} = \eta_i + \mu_{it} \qquad (9)$$

we can see a potential source of problems for Ms. Speedy's more clearly. Suppose η_i measures a household's taste for education. Higher η_i makes a household more likely to join the program, but also more likely to keep their children in school in any case. Thus η_{it} is not independent of P_{it}.

Two solutions using panel data

The problem is that participants and non-participants are not comparable because they have different (unobserved) tastes for education. Our goal is to take these unobserved effects out of the equation.

#1. Fixed effects estimation

$$
\begin{aligned}
S_{it} &= a + bP_{it} + cX_{it} + \varepsilon_{it} \\
\Rightarrow \overline{S}_i &= a + b\overline{P}_i + c\overline{X}_i + \overline{\varepsilon}_i \\
\Rightarrow S_{it} - \overline{S}_i &= a + b(P_{it} - \overline{P}_i) + c(X_{it} - \overline{X}_i) + \varepsilon_{it} - \overline{\varepsilon}_i
\end{aligned}
$$

where $\overline{X}_i = (\sum_{t=A,B} X_{it})/2$.

By removing each household's mean value (over time) from both sides of the equation, we remove the role of tastes and remove the bias they introduce.

Two solutions using panel data

#2. First differencing estimation

$$
\begin{aligned}
S_{it} &= a + bP_{it} + cX_{it} + \varepsilon_{it} \\
\Rightarrow S_{it} - S_{i,t-1} &= a + b(P_{it} - P_{i,t-1}) + c(X_{it} - X_{i,t-1}) + \varepsilon_{it} - \varepsilon_{i,t-1} \\
\Rightarrow \Delta S_{it} &= a + b\Delta P_{it} + c\Delta X_{it} + \Delta\varepsilon_{it}
\end{aligned}
$$

Note:

▶ Restricting our attention to changes removes the effect of time-invariant characteristics.

▶ With two periods, FE and FD are algebraically identical.

▶ With more than 2 periods, they may be different, which as important implications (see handout).

▶ NB: FE may be used in situations other than panels.
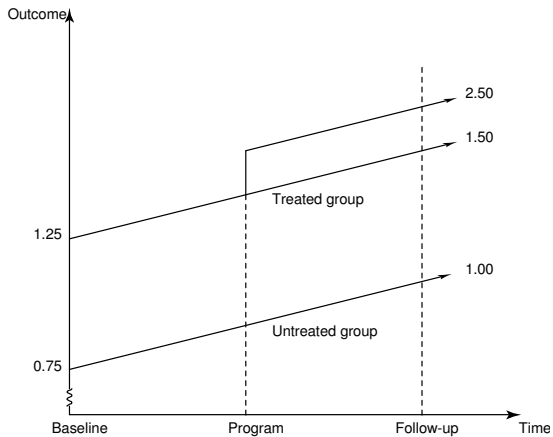
Diff-in-Diff and Natural Experiments

One particular application of FD estimation is the differences-in-differences approach to estimate treatment effects.

▶ You need data from a baseline survey (before the intervention or policy was introduced) and a follow-up survey (after the treatment) that spans treated and untreated groups.

▶ Calculate the mean difference or change in the outcome (say schooling) for the treatment and control groups.

▶ Calculate the difference between these two mean differences.

Diff-in-Diff and Natural Experiments

The DiD methodology presents a clear way to analyze the effects of policies implemented at different times in different places. Commonly applied to natural experiments.

- ► Galiani et al. (2005), "Water for Life"

- ► Besley and Burgess (2003) "Can Labor Regulation Hinder Economic Performance? Evidence from India"

- ► Levitt (1997) on the impact of police on crime rates

Unnatural Experiments

Besley and Case (2000) cast doubt on the use of DiD to evaluate policies without worrying more about our identification strategy.

▶ They review a large literature exploiting policy differences between states in the U.S. to do policy evaluation.

▶ They ask a simple, troubling question: why did policymakers in state A decide to raise labor standards while those in B did not?

▶ "If state policy making is purposeful action, responsive to economic and political conditions within the state, then it may be necessary to identify and control for the forces that lead policies to change if one wishes to obtain unbiased estimates of a policy's incidence."

Unnatural Experiments

So how do we get around endogenous placement of programs?

▶ Include characteristics used for targeting in the X vector:

$$\Delta S_{it} = a + b\Delta P_{it} + c\Delta X_{it} + \Delta\varepsilon_{it} \tag{10}$$

If we think these X factors affect changes in the outcome, not just its level, we can write

$$\Delta S_{it} = a + b\Delta P_{it} + c_1\Delta X_{it} + c_2 X_{i,t-1} + \Delta\varepsilon_{it} \tag{11}$$

▶ In addition to regression approaches, we can control for these X factors with matching techniques (next slide)

▶ Beyond that, we need to consider IV techniques, which we discuss tomorrow.

Overview of Propensity Score Matching

The idea of matching estimators is to compare each treated observation to non-treated observation that is as similar as possible.

Why match instead of using a regression?

▶ Good question. This isn't magic. You're still only controlling for observable characteristics.

▶ However, regressions require you to specify a functional form (income depends linearly on education for instance). Results may be fragile to functional form.

▶ The issue of 'common support': Regression tempts you to compare the incomparable. Suppose in reality PROSCOL raises schooling outcomes by 2 years. All non-poor households join PROSCOL and get 12 years schooling. All poor household don't participate and get 8 years schooling. With (near) perfect collinearity, how can we disentangle the effects of PROSCOL and income?

Treatment Effects by Propensity Score Matching

In the context of our problem, the general form of the treatment effect estimated by PSM would be

$$G = \frac{1}{N} \sum_{i \in \{P=1\}} \left(S_{1i} - \sum_{j \in \{P=0\}} \phi\left(i,j\right) S_{0,j} \right)$$

In words, the estimate of the treatment effect is the average distance (in terms of schooling) between each participant household and it's comparison group.

Treatment Effects by Propensity Score Matching

The comparison group for each data point is usually constructed in one of two ways (the difference is just in how we define the $\phi(.)$ function):

▶ Nearest neighbor matching: as its name implies, this approach compares each observation to the observation(s) with the closest p-score.

▶ Kernel matching: The idea behind kernel matching is to compare observation i in the program to the entire sample of non-participants, but to weight them by how similar they are to i. Thus we can write:

$$\phi(i,j) = \frac{K\left(p(x)_j - p(x)_i\right)}{\sum_{j=1}^{N_{C,i}} K\left(p(x)_j - p(x)_i\right)}$$

where K is some symmetric density function which has its maximum when its argument is zero.

Propensity Score Matching: Step by Step

1. Estimate the probability of participation using a probit or logit model. Create the predicted value or 'propensity' to participate for each observation, $\hat{p}$.

2. Now limit your sample to the area of common support. (If $\hat{p}$ ranges from .4 to 1 for participants, but from .1 to 1 for non-participants, you may want to exclude non-participants for whom $\hat{p}_{\mathsf{i}}.4$.)

3. Pick your comparison group for each observation. 'Nearest neighbor' matching, kernel matching, etc.

4. Measure the gap in outcomes between each observation and its comparison group. This is your measure of the treatment effect for that individual observation.

5. Calculate the mean of these differences.

6. Use bootstrapping to calculate standard errors.