Chapter 13 Pooling Cross Sections Across Time: Simple Panel Data Methods

Panel data looks at set of observations that have a cross sectional dimension and a time dimension.

Two types of data sets:

1. Independently pooled cross section
   Random sample from large population at different point in time
      Advantage: Balanced
      Disadvantage: Can't control for individual level differences

2. Panel data set or longitudinal data
   Follow same set of individuals, states, firms, families over time
      Advantage: Can control for individual differences
      Disadvantage: Sometimes need to deal with non-random attrition

"Big picture" for Chapters 13 and 14:

1. Panel data allows us to more closely replicate experimental design

What makes an experiment "ideal"?

--Control group and experimental group are identical (random assignment)
--individuals aren't choosing the "treatment"
--Observe change in behavior before and after "treatment"
-- Observe for long enough to see effects
-- Outcome of interest is measured correctly

→ Panel data is inherent in idea of ideal experiment
→ Observe at least 2 groups (experiment, control) at 2 or more points in time (before and after treatment)

Usually do not have an actual experiment, but a "natural experiment" or a "quasi-experiment"

Effective use of panel data often relies on determining relevant "treatment" and "control" groups

2. Panel data helps to resolve issues of "omitted variables"

Many economically important variables are unobserved. Unobserved ability, productivity, reservation price, reservation wage, etc.

Problem is that many times unobserved characteristics are correlated with the "treatment" (or other x variables) of interest.
Can use panel data methods to control for some types of omitted variables

13.1 Pooling Independent Cross Sections Across Time

Random sample drawn each time period→ Independently pooled cross section

Why use this type of data?

Increase sample size—gives more power to test due to (1) greater number of observations and (2) greater variation (remember, more variation in X reduces size of standard errors for estimates of beta)

Model specification issues:

1. Does mean of independent var change in each year?
   →Implies should include year dummy variables

   Different styles of notation:

   - $Y_{it} = \beta_0 + \beta_1 X_{it} + u_t + e_{it}$ (Error term has a component that is specific to each year t)
   - $Y_{it} = \beta_0 + \beta_1 X_{it} + 1996_t + e_{it}$ (Model includes a dummy variable for 1996)
   - $Y_{it} = \beta_0 + \beta_1 X_{it} + \delta D_t + e_{it}$ (Model includes vector of time dummy variables)

2. Do we expect the relationship between the dependent var and the independent var to be different in each year?
   →Implies should include interaction terms between independent variables and time dummy variable

   Note that full interaction is identical to estimating 2 separate equations

Which model should be used (interactions, structural breaks, etc)? Use Chow test to determine.

What is a Chow test?
F test to determine whether a multiple regression function differs across 2 groups

SSR from pooled regression is the restricted
SSR from 2 separately estimated time periods (full interaction) is unrestricted

$$F = \frac{[SSR_{pooled} - (SSR_1 + SSR_2)]}{SSR_1 + SSR_2} * \frac{n - 2(k+1)}{k+1}$$

Another way to test→do full interaction and then test whether year dummy and all interaction terms are jointly significant

13.1 Policy Analysis with Pooled Cross Sections—Difference in Difference Estimators

Difference-in-Difference estimators most closely replicate the "experimental design"

Like in ideal experiment,

Some event changes incentives, environment for a "treatment group"
Control group is not affected by policy change

Like in ideal experiment, treatment should be exogenous

is not determined by outcome of interest
is not correlated with unobserved characteristics of treatment and control group
is not self selected
often are policy changes implemented for different areas

However, unlike ideal experiment, in a natural experiment, treatment and control groups are usually not identical. Need to control for observable characteristics, have data from before and after treatment.

Basic D-D model

Two groups, two time periods

$y = \beta_0 + \delta_0 d2 + +\beta_1 dTR + \delta_1 d2{*}dTR$ + other factors +e

dTR is the dummy for Treatment group.

→Controls for permanent differences in average y for treatment and control

d2 is the dummy for post policy time period

→Controls for differences in average y over time common to both treatment and control

Interpretation of coefficients:

$\hat{\beta}_0$ average y in control group before change

$\hat{\beta}_1$ difference in average y between Treatment and Control before change

$\hat{\delta}_0$ change in average y for Control group over the 2 periods

$$\hat{\delta}_1 = (\bar{y}_{2,T} - \bar{y}_{1,T}) - (\bar{y}_{2,C} - \bar{y}_{1,C}) = (\bar{y}_{2,T} - \bar{y}_{2,C}) - (\bar{y}_{1,T} - \bar{y}_{1,C})$$

So δ_1 is our measure of the effect of the policy
=Average treatment effect

Known as the “Double difference estimator” or “Difference in difference estimator”

Note that δ_1 is NOT a marginal effect in the sense that it does not measure the effect on the marginal individual. Represents averages across groups. In that sense it is not a “behavioral” response— it averages effects for non-marginal individuals (where the behavioral response = 0) and marginal individuals (those with a behavioral response)

Key Assumption for Double-Difference Model:

Trend for control is what would have had in treatment group in absence of treatment

If treatment occurs in a group with a different trend, the D-D estimator will under/overstate the treatment effect

1. Is the “control” group an adequate comparison sample?
2. Is the trend for control the same as would have had for treatment?

Potential empirical approaches to address this issue:

1. **Show that pre-treatment means look similar**
   - Remember that D-in-D model does not require similar *means* but similar *trends*
   - However, if means are really different, have more concerns about whether control group is a good comparison group with similar trends.

2. **Pick narrow control group with similar pre-treatment trends**

3. If have more than 2 years before treatment**, Test with a “placebo treatment”**
   - Add “leads” to the model—treatment should not change outcomes *before* it appears
   - If it does, then have concern that trends and intervention covary

4. If have at least 3 time periods (really, better to have more than that) **Add group specific time trends** We’ll talk about this again in the next chapter.

13.2 Two Period Panel Data Analysis—First Differenced Model

Last section used 2 groups with a clear division between "treatment" and "control." Treatment was also assigned exogenously.

But what if data contains many groups and the "treatment" is something continuous rather than binary? What if "treatment" status is self selected?

Basic Two Period model:

$$y_{it} = \gamma_0 + \phi_0 d2_t + \gamma_1 x_{it} + a_i + u_{it}, \quad t = 1,2 \quad i = dummy\ for\ group$$

i denotes individual, producer, state, school district, etc.

x is key "treatment" variable of interest, but is no longer a dummy var, may not be exogenously assigned

Like before $d2_t$ is a dummy that equals 0 for period 1 (pre treat—before change in x), and 1 for period 2 (post treat—after a change in x)

$a_i + u_{it}$ Unobservables--2 types: time varying and constant.

a_i represents time constant (permanent, fixed) unobserved variables that are specific to group i.

Also called a FIXED EFFECT—because doesn't change over time. Also called UNOBSERVED HETEROGENEITY. Represents any permanent unobserved variables.

u_{it} captures any time varying error. Idiosyncratic error.

What if ignore fact that error is composed of these 2 parts?

Estimate $y_{it} = \gamma_0 + \phi_0 d2_t + \gamma_1 x_{it} + v_{it}$

What is problem here? Well, maybe nothing. When would there be a problem?

When v_{it} is correlated with x. Then have a specification error—conditional mean of true error term is not zero. Coefficient on x is biased.

One common source of bias--the permanent, unobserved characteristics that may be correlated with the x variable of interest.

BUT if a_i is constant over time, can do a transformation:

First Differenced Model:

$$y_{i2} = \gamma_0 + \phi_0 + \gamma_1 x_{i2} + a_i + u_{i2}$$

$$- \ y_{i1} = \gamma_0 + \gamma_1 x_{i1} + a_i + u_{i1}$$

→ $(y_{i2}\text{-}y_{i1}) = \phi_0 + \gamma_1(x_{i2}\text{-}x_{i1}) + (u_{i2}\text{-}u_{i1})$

Rewrite as $\Delta y_i = \phi_0 + \gamma_1 \Delta x_i + \Delta u_i$

We can estimate this "first differenced" equation directly using the transformed variables.

a_i has been removed from the model—Permanent unobserved characteristics have been "differenced" out

Now the key assumption is Δu_i is uncorrelated with Δx_i

How does First-Differenced Model relate to Double-Difference Model?

$\Delta y_i = \phi_0 + \gamma_1 \Delta x_i + \Delta u_i$

$y_i = \beta_0 + \delta_0 d2_t + +\beta_1 dTR_i + \delta_1 d2_t * dTR_i + e_{it}$

What if x is binary, as in D-D model?

Δx_i is 0 for groups that never change (control group) and 1 for groups that do change (treatment)

ϕ_0 represents the average change in y when Δx_i is zero (that is, the trend for the control group) = δ_0

$\phi_0 + \gamma_1$ represents the average change in y in the group where Δx_i is one (that is, the trend for the treatment group)= $\delta_0 + \delta_1$

γ_1 is therefore the difference in the difference (change in treatment relative to control)

The difference-in-difference model and the first-differenced model then will give identical results with two groups, two time periods, and a binary treatment.

The final sections of this chapter are straightforward generalizations of this material, and will not be covered in detail in lecture.

Policy Analysis with Two Period Panel Data

Note that before when we did policy analysis, I didn't observe the SAME woman in 2 different states. I had different women observed before and after a policy change. With panel data, however, I have the same district, person, etc. observed before and after a policy change.

This gives us more precise estimates—allows us to do differences to remove any time individual specific time constant differences. Otherwise, interpretation is the same.

T=dummy for treatment status

$$y_{it} = \beta_0 + \delta_0 d2_t + +\beta_1 T_{it} + a_i + u_{it}$$

$$\hat{\beta}_1 = \Delta\bar{y}_{treat} - \Delta\bar{y}_{control}$$

Differencing with More than 2 Time Periods

Not going to discuss this in detail—generalizes from what have said. With 2 years of data, start with 2 observations for each individual, district, firm, etc.
First difference reduces this to 1 observation.

With 3 years of data, do a first difference and reduces down to 2 obs—one that is change from year 1 to 2 and second that is change from year 2 to 3.

May need to deal with autocorrelation in this case. Book discusses this.

Also with multiple years, can do even more to control for unobservables.

Recall model $\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$

Key assumption is Δu_i is uncorrelated with Δx_i