Chapter 14 Advanced Panel Data Methods

$$y_{it} = \beta_1 x_{it} + complicatederrorterm, \quad t = 1,2,..T$$

β is interpreted to mean that an increase in x of one unit leads to a prediction, in all cases, that y will increase by **β** units.

The emphasis is on "in all cases":

In a panel, where does variation in X come from?
In other words, What variation identifies β?

I expect the same difference in y if

1. I observe two different subjects with a one-unit difference in x between them, and

2. I observe one subject whose x value increases by one unit.

For example, suppose y is income and x is "lives in the South of the United States":

1. If I compare two different people, one who lives in the East (x1=0) and another who lives in the South (x1=1), I expect the earnings of the person living in the South to be lower because, on average, all prices and wages are lower in the South. That is, I expect the coefficient on x1 will be less than 0.

2. On the other hand, if I observe a person living in the East (x1=0) who moves to the South (x1=1), I expect that the earnings increased, or why else would that person move? That is, I expect b1 will be greater than 0.

There are really two kinds of information in cross-sectional time-series data:

1. The cross-sectional information reflected in the changes **between** subjects

2. The time-series or within-subject information reflected in the changes **within** subjects

Deciding which specific panel data model adopt requires thinking about kind of variation in x to be used to IDENTIFY **β** —Ask what is the source of variation in my model that drives this effect?

Are these two sources of variation in X, within and between, likely to give the same effect on y?

14.1 Fixed Effects estimation

Suppose model is following:

(1) $y_{it} = \beta_1 x_{it} + a_i + u_{it}, \quad t = 1,2,...T$

As in chapter 13, the concern is that the estimate of β_1 will be biased if x is correlated with a_i—the fixed, unobserved characteristics.

First differencing was one way to eliminate these fixed unobserved components

An alternative (related) method is a FIXED EFFECT TRANSFORMATION

For each i, average over time for each individual
(2) $\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{u}_i$

Subtract (2) from (1)

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i, \quad t = 1,2,..T$$

That is, we regress the individual-demeaned y on individual-demeaned x

What kind of variation then is this using to identify β_1? The WITHIN variation—the variation in x over time for an individual. All average differences in ys or xs between individuals have been wiped out.
→Fixed effect estimator also called the WITHIN estimator.

Between estimator

This estimator is analogous, but here subtract the mean over time. Now β_1 is identified by variation in ys, xs, between individuals, not over time for same individual.

When would that be useful? When have something about time period that is specific that don't observe. Usually, the problem of individual specific error component is more common—that is, $\bar{x}_i$ is often correlated with a_i

Will come back to the between estimator though when talk about random effects—there the idea is that if DONT have problem that $\bar{x}_i$ is correlated with a_i then can use both sources of variation and get a more efficient estimator.

Dummy Variable Regression

One way to view this model is to think of a_i as a parameter to be estimated for each individual. Way we do this is to put in a dummy variation for each i (person, firm, state, etc—whatever the unit is that we observe over time).

What will be the value of the fixed effect? Mean for that group.

This give us EXACTLY the same estimates of the βs, their standard errors, etc. as using a demeaned transformation.

Fixed Effects or First Differencing?

In last chapter we also talked about differencing the data. That also dealt with unobserved effects. (Instead of subtracting the mean, we subtract one period from the other.)

What is the difference?

T=2—no difference in the estimated coefficients.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_{X^2}},$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

T=3+ The two methods will not give identical coefficients. However, both estimators are unbiased, consistent

Large N, small T

Which model choose depends on structure of errors over time—is there serial correlation in the u_{it} idiosyncratic errors?

No: fixed effects more efficient than first difference estimator

Yes: first differencing may be better—the Δu_{it} may have less autocorrelation

T large, N small

Fixed effect estimator--inference sensitive to violations of assumptions with small n
Use first differences—can appeal to CLT because of large T

Bottom line: Often want to check both and see if results are different—spend time looking at structure of errors over time if are—will discuss more in time-series chapters

What about missing data?

Often in panels, have an UNBALANCED panel—missing data on some individuals in some years. Dummy variable/fixed effect regression still works fine, although note that any individuals with only 1 observation get dropped.

If "attrition" or reason are missing is random—or at least uncorrelated with u_{it}, then not a problem. However, if IS related to u_{it}, then can lead to biased estimates. Will discuss models to deal with selection later.

Comparison with DD Model

Like with DD models, FE model control for time constant differences in means. FE models control for any permanent, unobserved variables

Like with DD models, are often concerned about differences in *trends* in unobserved variables.

Several ways to deal with that.

- Difference data over time a second time. This will subtract any unobserved/omitted variables that have a constant trend. See Hoxby paper for an example.

- Include interaction between individual fixed effect and a trend variable

  This is commonly used in DD style papers that use states or areas as the unit of observation

  $$y_{s,t} = \alpha + \beta X_{s,t} + \Theta_s \text{Time} + \xi_s + \eta_t + \varepsilon_{s,t}$$

  Note that Θ_s is a vector of s different coefficients on time—one for each state. This model still also includes state fixed effects and year fixed effects. In practice, sometimes papers will choose between state specific time trends and year fixed effects.

<u>Strengths and Weaknesses of First Differenced/Fixed Effect Models</u>

STRENGTHS: Controls for unobserved, time invariant effects that are correlated with error. A huge advantage when omitted variable bias is an issue.

WEAKNESSES:

1. <u>Amplify Measurement error in x</u>→ If x is not measured perfectly, have a noisy measure – a noisy measure→ much of the variation in Δx_i may be due to measurement error, rather than true underlying variation. (Amplifies the ratio of "noise" to "signal")

   What is effect of measurement error? Attenuated coefficient

   See proof for this—do with FD model

2. <u>Can't estimate effect of permanent characteristics.</u>

   In this model, can only obtain estimates of things with a Δx_i

   For example, being in the South may lead schools to have different lower test scores. Can't estimate this effect because there is no change over time: $\Delta x_i = 0$

   Similarly, can't include a dummy variable for each state and state level, permanent characteristics—why? Perfect collinearity.

   Often have to be careful about this in fixed effect models. "Dummy variable trap" can be harder to recognize on occasion. Check for dropped variables and identify reason drop.

3. <u>Less variation in differences</u>. Many times there is less variation in changes over time that there is across individuals in a cross section. More variation in levels of unionization across districts than variation in how much unionization changed.
   What is the consequence? Little variation in x→larger standard errors for β_1

4. <u>Source of variation for estimation is less clear</u>

   Fixed effect estimation removes some of the variation since subtracting the mean differences across unit of observation

14.2 Random Effects Models

Recall original model:

(1) $y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it}, \quad t = 1,2,..T$

Again use fixed effects if a_i is correlated with x_{itj}

In other words, going back to our within and between discussion, we thought that people who worked for non-profits were different from those that didn't. As a result, we wanted to look at the within- variation only—that means that our fixed effect estimator was identified by people who SWITCHED from one sector to the other. NOT by variation across individuals.

But what if we thought wasn't the case that a_i was correlated with Xitj

In that case want to use all the variation to get more efficient estimates.

Two ways to think about random effects models:

- Random effects model is a matrix weighted version of the between- and the within-(fixed effect) estimators.
- Random effects model is a GLS version of Pooled OLS model, accounting for fact that errors are serially correlated

Random effects model key assumption:

cov(x_{itj}, a_i) = 0, t=1, 2, ...T; j=1,2,…,k

Note that either using single cross section or pooled data will give us consistent estimates of betas.

However, doesn't exploit all the variation if use only cross section.

Not going to derive the random effects estimator.

But again, think about combining the variation **within** an individual over time and the variation **between** individuals at a point in time. How do we combine these two sources of variation? Weighted average. Weight is this:

$$\lambda = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}}$$

Recall FE estimator is this:

$$y_{it} - \bar{y}_i = \beta_0 + \beta_1(x_{it1} - \bar{x}_{i1}) + \ldots + \beta_k(x_{itk} - \bar{x}_{ik}) + u_{it} - \bar{u}_i, \quad t = 1,2,..T$$

R.E. estimator is this:

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1(x_{it1} - \lambda\bar{x}_{i1}) + \ldots + \beta_k(x_{itk} - \lambda\bar{x}_{ik}) + (v_{it} - \lambda\bar{v}_i)$$

Comparing RE estimates and FE estimates:

λ =1 for fixed effect. As quantity under square-root sign approaches zero→λ=1

- T is big→ lots of variation across time for each individual→ more like fixed effects
- σ^2_a is big→lots of variation in the "fixed effects"→ more like fixed effect estimate
- σ^2_u is small relative to σ^2_a→idiosyncratic variation is small—more of the variation is from fixed effect

Summary of RE:

- Random effects estimators are a weighted average of the between estimator (variation between individuals in a cross section) and the within/fixed effect estimator (variation within individuals over time)
- Random effects estimators will be **consistent and unbiased** ***if fixed effects are not correlated with x's***. Fixed effects estimators will always be consistent and unbiased (under usual GM assumptions)
- Random effects estimators will be more **efficient** (have smaller standard errors) than fixed effects estimators because they use more of the variation in X (specifically, they use the cross sectional/between variation)

14.3 Applying Panel Data methods to Other Data Structures

One common application is a "cluster sample" For example, may have a_i that are correlated within the same cluster (e.g., students in same school)
Solution is to use panel-robust/cluster/sandwich standard errors (different names)

Also, with fixed effect models, are not accounting to the potential autocorrelation within the errors for each i. Very common to also use clustered standard errors with fixed effect models.

Basic overview:

With random effects, use precise distributional assumptions

- within-school correlation takes the same form for all schools,
- each pupil within a school is correlated equally with any other pupil in the school.

With clustered standard errors don't model within school correlation explicitly

- allow for arbitrary correlation within schools,
- the form of this correlation can vary from school to school.

Trade offs:

- Random effect give more efficient estimates if modelling of the correlation caused by clustering is correct. If it isn't, coeffs and SEs are wrong. (recall RE relies on zero covariance between ai and xit's)

- With clustered standard errors, get consistent estimates across a broad range of possible forms of the correlation, but they won't be as efficient when you know the exact form.

In STATA, **`reg yvar xvar, cluster(school)`**

Panel Robust Sandwich Standard Errors/Cluster-Robust standard errors

Version of robust standard errors in a panel context

In a panel context, very likely that Cov[uit, uis]>0 for t $\neq$s.
If ignore this serial correlation, will greatly underestimate standard errors and overestimate t-stats

May have already done some panel transformation (f.d., f.e., r.e.) to get the transformed variables:

$$\tilde{y}_{it} = \tilde{\boldsymbol{x}}_{it}\beta + \tilde{u}_{it}$$

Notation:
Stacking observations over time periods for a given individual

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{x}}_i\beta + \tilde{\boldsymbol{u}}_i$$

Then stack by the N individuals

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{x}} \quad \beta + \tilde{\boldsymbol{u}}$$

Three different ways we can write the OLS estimator:

$$\hat{\beta} = [\tilde{X}'\tilde{X}]^{-1}\tilde{X}'\tilde{y} = \sum_{i=1}^{N}[\tilde{X}_i{}'\tilde{X}_i]^{-1}\tilde{X}_i{}'\tilde{y}_i = \sum_{i=1}^{N}\sum_{t=1}^{T}[\widetilde{X_{it}}{}'\widetilde{X_{it}}]^{-1}\tilde{X}_{it}{}'\widetilde{y_{it}}$$

Most convenient to use varies with contest

When with this estimator be consistent?

$$\hat{\beta} = \beta + \sum_{i=1}^{N}[\tilde{X}_i{}'\tilde{X}_i]^{-1}\tilde{X}_i{}'\tilde{u}_i$$

Assume independence over I. Now the exogeneity condition for consistency is

$E[\tilde{X}_i{}'\tilde{u}_i]$=0

This is a stronger assumption than $E[\tilde{X}_{it}{}'\widetilde{u_{it}}]$=0

(Discussed this with serial correlation?)

The panel-robust estimate of the asymptotic variance matrix (that is, one that controls for both autocorrelation and heteroskedasticity) is analogous to the usual robust standard errors:

$$\widehat{V[\tilde{\beta}]} = \left[\sum_{i=1}^{N}[\tilde{X}_i{}'\tilde{X}_i]\right]^{-1}\sum_{i=1}^{N}[\tilde{X}_i{}'\hat{u}_i\hat{u}_i{}'\tilde{X}_i]\left[\sum_{i=1}^{N}[\tilde{X}_i{}'\tilde{X}_i]\right]^{-1}$$

where the $\hat{u}_i$ are the usual predicted errors

Again, this assumes independence over i, but permits V[uit] and Cov[uit, uis] to vary with i,t,s

An equivalent expression is

$$\widehat{V[\tilde{\beta}]} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}[\widetilde{X_{it}}{}'\widetilde{X_{it}}]\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}[\widetilde{X_{it}}\widetilde{X_{is}}{}'\hat{u}_{it}\hat{u}_{is}]\left[\sum_{i=1}^{N}\sum_{t=1}^{T}[\widetilde{X_{it}}{}'\widetilde{X_{it}}]\right]^{-1}$$

So if use traditional robust standard errors, will adjust for heteroskedasticty, but not autocorrelation. In panel setting, problem of autocorrelation is much bigger.

Summary:

	Pooled OLS is true model $y_{it} = \beta_1 x_{it} + u_{it}$	Random Effects is true model $y_{it} = \beta_1 x_{it} + a_i + u_{it}$ a_i not correlated with x_i	Fixed Effects is true model $y_{it} = \beta_1 x_{it} + a_i + u_{it}$ a_i correlated with x_i
Pooled OLS	Consistent Efficient	Consistent	Inconsistent
Within (Fixed Effect)	Consistent	Consistent	Consistent Usually most efficient, especially with large N, small T—likely to require clustered errors to get correct se
First Difference	Consistent	Consistent	Consistent Can be more efficient than F.E. with small N, large T, depending on structure of autocorrelation
Random Effects	Consistent	Consistent Efficient	Inconsistent

Testing for Which Model to Use

1. Choosing Between Fixed Effects and Pooled OLS

Is it appropriate to include set of dummy variables?

- **Use F-test** to decide, just as would with any set of variables

2. Compare Between Random Effects and Pooled OLS

If use OLS on pooled data, have T*n observations—can overstate precision of estimates—really have same n people observed T times. Need to account for serial correlation in errors

Let Vit = ai+uit—person specific error and time specific error

Look at correlation in this over time for same individual:

corr(vit, vis) = $\sigma^2_a/(\sigma^2_a+\sigma^2_u)$

This correlation over time can be substantial→ OLS stats for pooled model ignore this.
→Need a GLS estimator (remind me what this is?)

- **Breusch Pagan LM** test for Unobserved heterogeneity

Is there unobserved heterogeneity? Ho: $\sigma^2_a = 0$

NOTE: This is NOT asking if the person specific error (the unobserved heterogeneity) is correlated with the Xs. This is only asking IF IT EXISTS. How do we know it exists? Because the errors for an individual are related.

One way to approach—test for serial correlation in the vi = ai + uit

or we can do a test directly based on σ^2_a

in Stata, do **xttest0**

3. <u>Choosing Between Random Effects and Fixed Effects</u>

Why might we consider random effects instead of fixed effects?

1. When are looking at a variable that doesn't vary over time—can't use a fixed effects estimate.

   For example, if were looking at effect of gender on wages, couldn't use a fixed effect model. Could include random effects if had a panel. Better than OLS, although note that if key variable of interest is not determined exogenously, will still have biased coefficients

2. When are sure that x is exogenously determined—say by an experiment

3. If Hausman test indicates that rejection of fixed effects is appropriate—*but be careful*

Hausman test

This is a test we will see again in IV context—the structure is the same

Need two estimators to perform a Hausman test

1. An estimator (R.E.) that is consistent AND efficient under HO (no correlation) and inconsistent under HA
2. An estimator that is consistent under both (F.E.)

Basically look at how different the two estimates are—if they are really different, likely FE is more appropriate.

Test statistic:

$$(\hat{\beta}_{FE} - \hat{\beta}_{RE})'[V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \sim \chi_k^2$$

If have just one X, this is just a t-test for the difference in the coefficients

$$H = \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right) / \sqrt{(se\hat{\beta}_{FE})^2 - (se\hat{\beta}_{RE})^2}$$

Problem is may not be "really different" because even though point estimates are pretty different if sample variation in FE is so big can't say are statistically different. But that's not very satisfying—better to have a good argument based on theory for why aren't worried about correlation between xs and ai, and then use Hausman test to back you up.

Note: this is NOT about whether ai is "fixed" or random. It is about whether ai is correlated with xitj