LECTURE 11: FIXED EFFECTS

March 11, 2010

Plan for Today

- Random and Fixed Effects Estimation
- Raymond & Hanushek reanalysis
- Computer Assignment #1 and #2

Difference-in-Differences Estimator

- We want to find the effect of $\bar{Y}_T^{After} - \bar{Y}_T^{Before}$
- Some of this effect will be because of the program, but some will occur naturally over time
- The effect of time can be measured using the control group $\bar{Y}_C^{After} - \bar{Y}_C^{Before}$

$$\text{Impact of program} = \left(\bar{Y}_T^{After} - \bar{Y}_T^{Before}\right) - \left(\bar{Y}_C^{After} - \bar{Y}_C^{Before}\right)$$

$$(treatment + time) - \quad (time) \quad = treatment$$

Natural Experiments

- ***Natural Experiment***- An external agency other than the experimenter assigns subjects exogenously to treatment conditions
- Considerations:
  - Is your assignment to treatment and control truly exogenous?
  - Could the disruption (or shock) have been provoked as a result of underlying changes in the populations of participants themselves?
  - How wide should the "data-window" be on each side of the disruption?

Nesting within units

- **In social data, some "units" are often nested within other "units," in a hierarchical or multilevel structure**
  - Kids are nested within teachers or classes
  - Teachers and classes are nested within schools
  - Schools are nested within districts
  - Districts are nested within States
- **You have to respect these hierarchies in your data analysis if you want your estimation to be correct**
  - You must specify your regression models to account for the multilevel structure of the data.

Fixed Effects

- Helps to eliminate bias by controlling entirely for all observed and unobserved effects at some specified "level" of analysis.
- In reality, students within a school share **common unobserved experiences** that have the potential to make them behave similarly to each other
- BUT the OLS regression model assumes that each participant's residuals are independent of each other. Failure to accounts for this leads to:
  - Mis-estimated standard errors
  - Incorrect statistical inference

Fixed Effects

- Accounts for the <u>fixed, unobserved differences</u> between participants and non-participants that may exist, which are also correlated with the outcome

$$y_{it} = \beta_0 + \beta_1 TREAT_i + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \ldots + \gamma_k x_{ki} + \beta_2 Z_i + \varepsilon_{it}$$

- Say Z is unobserved motivation. If we think motivation is stable over time, and we have a measure of the outcome for each individual at more than two time periods, then we can account for each student's unobserved level of motivation.

Fixed Effects

- We use the fixed effects estimator to account for the effect of *any* unobserved characteristics that are stable over time
- With multiple observations per person, we can modify our regression model:

$$y_{it} = \beta_1 TREAT_i + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \ldots + \gamma_k x_{ki} + \alpha_i + \varepsilon_{it}$$

- Where:

$$\alpha_i = \beta_0 + \beta_2 Z_i$$

School Random Effects

Outcome, with subscripts *at two levels:*

- i = school; j = student

- Because they're in the structural part of the model, they're referred to as the "*fixed effects*" in the model

$$\boxed{OUTCOME_{ij}} = \beta_0 + \boxed{\beta_1} X_{ij} + \boxed{\beta_2} X_i + (\varepsilon_{ij} + u_i)$$

2 Residuals in the multilevel model, one at each level

Random effect of student, ε_{ij} :

- Student-level residual for student j in school i
- Same as regular regression residual.
- Represents the *unique unobserved effect of each student*, and is assumed independent across every student in every school.

Random effect of school, u_i :

- New *school-level* residual
- Value is identical for all students *within* a given school
- Responsible for providing the *unobserved effect of school* that is shared by all students within a school

Random Effects in STATA

```
xtreg outcome female enrol pfrl black hisp, re i(schid)
```

XTREG fits this type of multilevel model

Requests random effects model

Students are clustered *within school*, as defined by *schid*

```
xtreg outcome female enrol pfrl black hisp pctother pctother2, re i(schid)

Random-effects ML regression                    Number of obs      =      5131
Group variable (i): schid                       Number of groups   =        56

Random effects u_i ~ Gaussian                   Obs per group: min =         2
                                                               avg =      91.6
                                                               max =       314

                                                LR chi2(7)         =     84.35
Log likelihood  = -14416.951                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
    livework |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |    .756667   .1135648     6.66   0.000      .534084      .97925
       enrol |  -.0002599   .0001001    -2.60   0.009     -.000456   -.0000637
        pfrl |  -.0139089   .0077714    -1.79   0.073    -.0291406    .0013228
       black |  -.0051021    .187619    -0.03   0.978    -.3728286    .3626243
        hisp |    .749619   .1791884     4.18   0.000     .3984162    1.100822
    pctother |   .0324187   .0160925     2.01   0.044      .000878    .0639593
   pctother2 |  -.0002642   .0001475    -1.79   0.073    -.0005533    .0000249
       _cons |    18.5209   .6200724    29.87   0.000     17.30558    19.73622
-------------+----------------------------------------------------------------
    /sigma_u |   .5798191   .1019724     5.69   0.000     .3799569    .7796813
    /sigma_e |   3.998067    .039686   100.74   0.000     3.920284     4.07585
-------------+----------------------------------------------------------------
         rho |    .020599    .007127                      .0100921    .0391905
------------------------------------------------------------------------------
```

Estimated *variance components* are as before, but they are standard deviations and you must square them to obtain variances.

From Random to Fixed Effects

Now, let's move the *school-level random effect* into the "fixed" part of the model

$$OUTCOME_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 X_i + (\varepsilon_{ij} \boxed{+ u_i})$$

$$OUTCOME_{ij} = \beta_0 \boxed{+ u_i +} \beta_1 X_{ij} + \beta_2 X_i + \varepsilon_{ij}$$

- Now, instead of treating the unobserved "school-specific parts" (the u_i) as unknowns and part of the random effects, we have included the u_i in the "fixed" part of the multilevel model and we can try to estimate the magnitude and direction of each.
- This is easy using a dummy variable specification for school

From Random to Fixed Effects

$$OUTCOME_{ij} = \beta_0 + (\gamma_2 S_2 + \gamma_3 S_3 + \cdots + \gamma_{56} S_{56}) + \beta_1 X_{ij} + \beta_2 X_i + \varepsilon_{ij}$$

Where S= School

Now, we are treating the unobserved effects of school as *fixed-effects*, and we can fit the new *fixed effects of school multilevel model* using *OLS*, because our re-specification has returned the residual to its usual "independent across kids" form …

“Fixed-Effects of school” multilevel model

- Create a set of dummy variables to distinguish the *56* schools
- Include the dummies (minus a reference school) as “fixed effects” (i.e., predictors) in a regular *OLS* model.

Create the school dummies

Include the school dummies (except for reference school, #1) as “fixed effects” (i.e. predictors)

Regular *OLS* regression routine

```
quietly tab schid, gen(school)
reg livework female black hisp pctother pctother2 school2-school56

      Source |       SS       df       MS              Number of obs =
-------------+------------------------------           F( 60,  5070) =
       Model |  4859.64516    60  80.9940861           Prob > F      =  0.0000
    Residual |          45  5070  15.9777622           R-squared     =  0.0566
-------------+------------------------------           Adj R-squared =  0.0454
       Total |          96  5130  16.7381871           Root MSE      =  3.9972

------------------------------------------------------------------------------
    livework |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .7361477   .1139796
       enrol |  (dropped)
        pfrl |  (dropped)
       black |   .1140067    .199588
        hisp |   .8087041   .1916357
    pctother |   .0202195   .0291537
   pctother2 |  -.0001618   .0002581
     school2 |  -.2866461   1.562476
     school3 |   1.479005   .6641335
        :    |      :          :
    school55 |  -1.254927   .8178375
    school56 |   1.469761   .7119348
       _cons |   18.06612   .8323094    21.71   0.000     16.43443    19.69781
------------------------------------------------------------------------------
```

$$School\_1: \quad \hat{\beta}_0 + \hat{u}_1 = 18.0661$$

$$School\_2: \quad \hat{\beta}_0 + \hat{u}_2 = 18.0661 - 0.2866 = 17.7795$$

$$School\_3: \quad \hat{\beta}_0 + \hat{u}_3 = 18.0661 + 1.4790 = 19.5451$$

$$\vdots$$

$$School\_13: \quad \hat{\beta}_0 + \hat{u}_{13} = 18.0661 - 5.3838 = 12.6823$$

The estimates can be interpreted as usual:

- For each predictor, X, the associated parameter estimates the *within-school* Y/X slope, *pooling across all schools.*

```
quietly tab schid, gen(school)
reg livework female black hisp pct

      Source |       SS       df
-------------+----------------------
       Model |  4859.64516    60
    Residual |  81007.2545  5070
-------------+------------------------------
       Total |  85866.8996  5130  16.7381871

                                               Adj R-squared =  0.0454
                                               Root MSE      =  3.9972

------------------------------
    livework |      Coef.    St
-------------+----------------
      female |   .7361477    .1
       enrol |  (dropped)
        pfrl |  (dropped)
       black |   .1140067    .
        hisp |   .8087041    .1
    pctother |   .0202195    .0
   pctother2 |  -.0001618    .0
     school2 |  -.2866461    1.
     school3 |   1.479005    .6
          :          :
    school55 |  -1.254927    .8
    school56 |   1.469761    .7
       _cons |   18.06612    .8
------------------------------
```

- Now that the fixed effects of school are explicitly included in the model, other school-level predictors *enrol* and *pfrl* have "dropped" out of the model.
- You *can no longer include any other school-level effects*. Why? Because the fixed effects of school *explain all possible school-level differences in the outcome*, including:
  - Those predicted by *observed* school characteristics, like *enrol* and *pfrl*.
  - Those predicted by school characteristics that are currently *unobserved!*

STATA XTREG lets you fit a “fixed effects” model *directly* using the “*fe*” option

```
xtreg livework female enrol pfrl black hisp pctother pctother2, fe i(schid)

Fixed-effects (within) regression               Number of obs      =     5
Group variable (i): schid                       Number of groups   =

                                                Obs per group: min =
                                                               avg =     9
                                                               max =

                                                F(5,5070)          =     13.40
                                                Prob > F           =    0.0000

------------------------------------------------------------------------------
    livework |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .7361477   .1139796     6.46   0.000     .5126985    .9595969
       enrol |  (dropped)
        pfrl |  (dropped)
       black |   .1140067    .199588     0.57   0.568    -.2772719    .5052853
        hisp |
    pctother
   pctother2
       _cons
-------------
F test that
```

Requests a *fixed effects of school* analysis, you don’t need to create the dummies yourself

Although the “fe” option ensures that the 56 fixed effects of school get included in the model, their estimates are not listed with the other parameter estimates in the table to conserve space

Considerations when using fixed effects

- Including a set of dummy variables in the model effectively allows each entity in the dataset to have his or her own intercept in the regression model.
- This intercept eliminates the effects of any and all characteristics of that are (1) unique to that student, (2) that remain fixed over time, and (3) that are related to *y*.
- We lose degrees of freedom (because we now have n-1 additional independent variables in our model) and therefore statistical power, but we gain unbiased estimates *as long as the unobserved differences between program participants and non-participants do not vary over time.*
- The treatment effects estimated using fixed effects models only have external validity for the population of students that have over-time variation in treatment status.

Example from last time: Traffic deaths and alcohol taxes

Unit of observation: a year in a U.S. state

- 48 (contiguous) U.S. states, so $n = 48$
- 7 years (1982,…, 1988), so $T = 7$
- Outcome variable: Traffic fatality rate (# traffic deaths in that state in that year, per 10,000 state residents)
- Independent variable: Tax on a case of beer

Panel Data with 3+ Time Periods: Fixed Effects Estimation

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it},\ i = 1,\ldots,n,\ T = 1,\ldots,T$$

- Suppose we had data from only three states: California, Texas, and Massachusetts
- Regression for 7 years of California data (that is, i = CA):

$$\begin{aligned} Y_{CA,t} &= \beta_0 + \beta_1 X_{CA,t} + \beta_2 Z_{CA} + u_{CA,t} \\ &= (\beta_0 + \beta_2 Z_{CA}) + \beta_1 X_{CA,t} + u_{CA,t} \end{aligned}$$

or

$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$

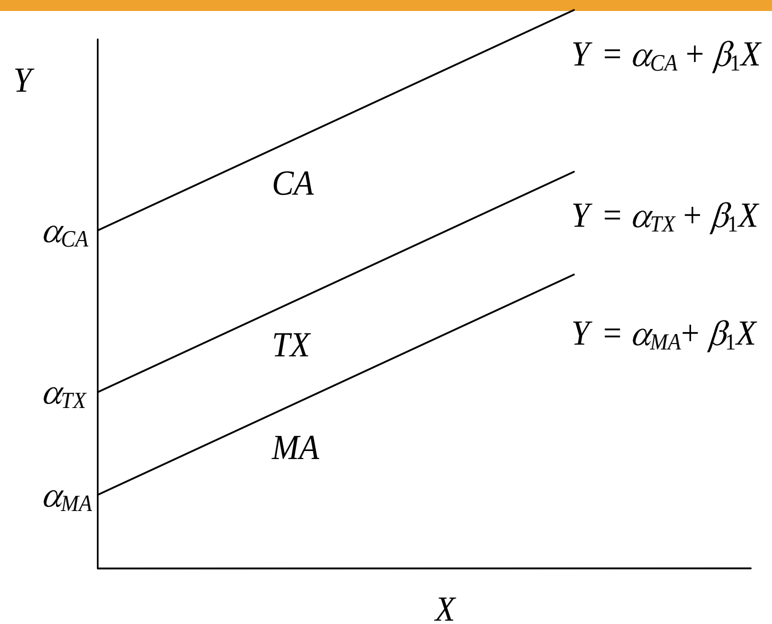
Regression lines for all 3 states

$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$

$$Y_{TX,t} = \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t}$$

$$Y_{MA,t} = \alpha_{MA} + \beta_1 X_{MA,t} + u_{MA,t}$$

The regression lines for each state



$$Y_{it} = \beta_0 + \gamma_{CA}DCA_i + \gamma_{TX}DTX_i + \beta_1 X_{it} + u_{it}$$

- DCA_i = 1 if state is *CA*, = 0 otherwise
- DTX_t = 1 if state is *TX*, = 0 otherwise
- Leave out DMA_i (*why*?)

Two ways to write fixed effects models

Dummy Variables form:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \ldots + \gamma_n Dn_i + u_{it}$$

where $D2_i = \begin{cases} 1 \text{ for } i=2 \text{ (state \#2)} \\ 0 \text{ otherwise} \end{cases}$, etc.

"Fixed effects" form:

$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}$$

- α_i is called a "state fixed effect" or "state effect" – it is the constant (fixed) effect of being in state i

Adding another set of fixed effects

- Other plausible omitted variables might vary over time but not across states:
  - e.g. safer cars; changes in national laws; changes in oil prices
  - Let these changes ("safer cars") be denoted by S_t, which changes over time but not across states.
- The appropriate panel data regression model is now:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + \beta_3 S_t + u_{it}$$

- Solution? Include a fixed effect for each year

What about the standard errors?

- OLS assumes that (conditional on X), the error terms for a given state are uncorrelated over time.
  - For example, $u_{CA,1982}$ and $u_{CA,1983}$ are uncorrelated
- Is this plausible? What enters the error term?
  - The effects of a snowy winter, the opening of a new divided highway, fluctuations in traffic density, etc.
- How likely is it that <u>all</u> of these omitted factors are uncorrelated over time within states?
  - Not very!

Clustered standard errors

- We often need to acknowledge that grouped observations <u>are not independent</u> and therefore provide less information than they would otherwise
- We do so by estimating OLS with <u>clustered</u> standard errors, which allow for the fact that the data contain "clusters" within which the error term is possibly correlated but outside of which (across groups) it is not.

Clustered Standard Errors

The solution is …

- Modify the regression model to represent the new reality
- Specify the regression model so that it allows each child's (or entity's) residual to be correlated with the residuals of other children in the school (so that the residuals are "correlated across children within-school").

Fixed Effects Regression Results
Dependent variable: *Fatality rate*

```
xtreg vfrall beertax y83 y84 y85 y86 y87 y88, fe i(state)
cluster (state)
```

	(1)	(2)	(3)	(4)
BeerTax	-.656** (.203)	-.656^{+} (.315)	-.640* (.255)	-.640^{+} (.386)
State effects?	Yes	Yes	Yes	Yes
Time effects?	No	No	Yes	Yes
Clustered SEs?	No	Yes	No	Yes

Significant at the **1% *5% $^{+}$15% level

Clarification: "Fixed Effects" Estimation in Cross-Sectional Data

- We can also use a set of dummy variables to eliminate sources of omitted variables bias in cross-sectional data
  - Example: Including "region fixed effects" in Fryer and Levitt's (2005) analysis of the test score gap

Random Effects

- Model the unobserved impact of the schools – the u_i – as *random effects*, by treating them as part of the random (residual) part of the regression model
- Convenient to implement
- Cheap -- it costs you only *one degree of freedom*, for a parameter to represent the variance of the new random effect, σ_u^2.
- Only a part of the school-level outcome variability is "explained" subsequently by school-level predictors that you have explicitly included in the model.

Fixed Effects

Model the school effects – the u_i – as *fixed effects*, by treating them as predictors in the fixed (structural) part of the regression model.

- Somewhat cumbersome, especially when you must create a large # of dummies and insert in an *OLS* model
- Expensive – it costs you *many degrees of freedom*, one for *each* of the fixed effects introduced into the model (here, that's the number of schools minus 1)
- Restrictive, once you put the fixed effects in the model, you can no longer introduce other interesting school-level predictors
- All school-level outcome variability is "explained" by the school fixed-effects – they represent all observed and unobserved school-level effects

Hanushek and Raymond (2003)

- "Make-or-break Exams Grow, but Big Study Doubts Value"

  – New York Times, December 2002

- Method of the Amrein and Berliner study?

- What is the fatal flaw?

Rerunning the Amrein-Berliner Data (Table 1)

When the actual test scores in the states Audrey Amrein and David Berliner identified as "high stakes" are compared with those in states without accountability systems, the high-stakes states show much more improvement.

	Increase in NAEP 4th-grade math scores		Increase in NAEP 8th-grade math scores	
	1992-2000	1996-2000	1992-2000	1996-2000
High-stakes states	9.2	4.2	8.8	4.5
No accountability states	3.8	2.3	4.0	1.7
High-stakes advantage	**5.3 points***	**1.9 points***	**4.8 points***	**2.8 points***
High-stakes advantage after adjusting for changes in students excluded from NAEP	5.2 points*	2.3 points*	3.7 points*	2.5 points*

* statistically significant at the .05 level

SOURCE: Authors

Identifying accountabilty effects

- Cross-sectional approach:

$$O_{st} = \beta_0 + \beta_X X_{st} + \beta_R R_{st} + \gamma A_{st} + (\rho_s + \varepsilon_{st})$$

- Diffs-in-diffs approach:

$$\underset{t,t^*}{\Delta} O_s = \beta_X \Delta X_s + \beta_R \Delta R_s + \gamma \Delta A_s + \Delta \varepsilon_s$$

- State fixed effects approach:

$$\underset{t,t^*}{\Delta} O_s = \beta_X \Delta X_s + \beta_R \Delta R_s + \gamma \Delta A_s + \delta_s + \Delta \varepsilon_s$$