Review Session 7

Fixed Effects

Ben Berger

March 8, 2023

Fixed Effects

$$\text{CIA: } \{Y_{0i}, Y_{1i}\} \perp D_i | W_i$$

Don't always observe all relevant variables W_i, biasing estimate. Can control for some types of unobservable variables using fixed effects.

Set Up/What You Need

- Panel data – data where the same individuals/groups/firms/units/etc. are observed multiple times.
  - Often, an individual observed in multiple time periods.
  - Sometimes, an individual nested within a larger unit (family, firm, school).

Basic Idea

- Control for all variation across units by removing cross-unit variation via demeaning or dummy variables.
- Why do we need panel data for this? If we don't have multiple observations per unit, then fixed effects control for all variation in data.

Fixed Effects

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

Panel data example, indexed by individual i and time t.

Individual-level fixed effects

- Control for characteristics that are fixed for each individual.

Time-level fixed effects

- Control for characteristics that are fixed in every time period.

Individual-level FEs and time-level FEs together

- Do control for characteristics that are fixed over time
- Do control for characteristics that are fixed for each individual
- Do not control for characteristics that vary over time *and* vary across groups.

Fixed Effects

Fixed Effects' Costs and Benefits

- Pro: Remove the OVB due to the observable *and* unobservable heterogeneity across observations that do not change over time.
- Con: Demeaning throws away all variation *between* units from our data

Assumption Needed for Causal Interpretation

- Treatment assignment is as good as random conditional on characteristics that are fixed.
  - Weaker assumption than OLS.
  - But for many applications this is still a strong assumption.
  - Can add also add controls to regression if they aren't collinear with fixed effects.

FE Exercise 1: Watch the Index

Consider the following specification for arbitrary dependent variable Y and independent variable X:

$$Revenue_{it} = \beta_0 + \beta_1 R\&D_{it} + \alpha_i + \lambda_t + u_{it}$$

for firm i, in year t with error term u_{it}. Our data includes 500 firms each observed in 10 years (i.e. it's a balanced panel).

Questions

1. Which fixed effects are specified in this model? Firm (i) and year (t) fixed effects
2. How many dummy variables are needed to include both firm and year fixed effects? $(500 - 1) + (10 - 1) = 508$
3. Could we add state fixed effects? If firms move states in our sample, then yes. Otherwise, no.

Estimating Fixed Effects Model in R

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \alpha_i + \lambda_t + u_{it}$$

Two equivalent approaches using `feols()`:

- Estimate model and report coefficients on all dummy variables by using the `i()` function to make dummy variables.
  - `feols(y ~ x1 + x2 + i(i) + i(t), data, cluster = ~i)`
- Estimate model without reporting coefficients on dummy variables.
  - `feols(y ~ x1 + x2 | i + t, data, cluster = ~i)`

Why cluster at the unit-level rather than time? Generally quite concerned that errors are correlated within a unit across time periods. Clustering by time period allows for covariance between all individuals within a time period, but no persistence of those correlations over time.

FE Exercise 2: Investment

Grunfeld's investment data is a *balanced* panel of 10 firms from 1935 to 1954.

The data contains data on the firms' annual investment, value, and capital stock. We can potentially use it to understand the determinants of firm investment. We will contrast OLS estimates to fixed-effects.

Run `install.packages("plm")` and confirm that the package loads properly.

FE Exercise 2: Investment

```
library(fixest)
data("Grunfeld", package = "plm")
head(Grunfeld)
```

```
  firm year   inv  value capital
1    1 1935 317.6 3078.5     2.8
2    1 1936 391.8 4661.7    52.6
3    1 1937 410.6 5387.1   156.9
4    1 1938 257.7 2792.2   209.2
5    1 1939 330.8 4313.2   203.4
6    1 1940 461.2 4643.9   207.2
```

FE Exercise 2: Investment

Consider the following models:

$$Investment_{it} = \beta_0 + \beta_1 Value_{it} + \beta_2 Capital_{it} + u_{it} \quad (1)$$
$$Investment_{it} = \beta_0 + \beta_1 Value_{it} + \beta_2 Capital_{it} + \alpha_i + u_{it} \quad (2)$$
$$Investment_{it} = \beta_0 + \beta_1 Value_{it} + \beta_2 Capital_{it} + \alpha_i + \lambda_t + u_{it} \quad (3)$$

Using `feols()` estimate:

- Model 1 using regular OLS
- Model 2 using OLS with firm-level dummy variables
- Model 2 using a fixed-effects specification
- Model 3 using a two-way fixed-effects specification

In all models, estimate standard errors clustered at the firm level.

FE Exercise 2: Investment

```
# OLS (pooled)
ols <- feols(inv ~ value + capital,
             data = Grunfeld,
             cluster = ~firm)
```

```
# Least Squares Dummy Variable estimator
lsdv <- feols(inv ~ value + capital + i(firm),
              data = Grunfeld,
              cluster = ~firm)
```

FE Exercise 2: Investment

```
# Fixed effects estimator
fe <- feols(inv ~ value + capital | firm,
            data = Grunfeld,
            cluster = ~firm)
```

```
# Two-way fixed effects estimator
twfe <- feols(inv ~ value + capital | firm + year,
              data = Grunfeld,
              cluster = ~firm)
```

FE Exercise 2: Investment

Dependent Variable:	Investment			
Model:	(1)	(2)	(3)	(4)
Variables				
Value	0.12*** (0.02)	0.11*** (0.02)	0.11*** (0.01)	0.12*** (0.01)
Capital	0.23** (0.09)	0.31*** (0.05)	0.31*** (0.05)	0.36*** (0.05)
firm = 2		172.2*** (50.3)		
firm = 3		-165.3*** (46.3)		
firm = 4		42.5 (76.8)		
firm = 5		-44.3 (69.3)		
firm = 6		47.1 (81.6)		
firm = 7		3.7 (77.0)		
firm = 8		12.8 (78.7)		
firm = 9		-16.9 (74.9)		
firm = 10		63.7 (91.0)		
Constant	-42.7* (20.4)	-70.3 (92.4)		
Fixed-effects				
firm			Yes	Yes
year				Yes
Fit statistics				
Observations	200	200	200	200

Clustered (firm) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

FE Exercise 2: Investment

Interpret $\hat{\beta}_{\text{capital}}$ in each specification. Note that investment, capital, and value are all measured in dollar terms.

- Regression 1
  - A \$1 increase in capital is associated with a \$0.23 increase in investment controlling for firm value.
- Regression 2
  - A \$1 increase in capital is associated with a \$0.31 increase in investment controlling for firm value and firm fixed effects.
- Regression 3
  - Same as Regression 2.
- Regression 4
  - A \$1 increase in capital is associated with a \$0.36 increase in investment controlling for firm value, firm fixed effects, and year fixed effects.

FE Exercise 2: Investment

- When adding firm fixed effects, what happens to:
  - Coefficient estimates?
    - $\hat{\beta}_{\text{value}}$ nearly unchanged
    - $\hat{\beta}_{\text{capital}}$ increases substantially
    - Capital negatively correlated with time-invariant factors that are positively correlated with investment
  - Standard errors?
    - FE explain residual variation in investment ($\downarrow$S.E.)
    - FE explain all variation across firms in explanatory variables ($\uparrow$S.E.)
    - Net effect is $se(\hat{\beta}_{\text{value}})$ stays about the same and $se(\hat{\beta}_{\text{capital}})$ decreases.
- What is the difference between the columns 2 and 3?
  - $\hat{\beta}_{\text{value}}$ and $\hat{\beta}_{\text{capital}}$ are the same, but we also estimate a bunch of coefficients on dummy variables.
  - Why would you want to do one versus the other?
    - If you are interested in the fixed effects themselves, then estimate them using dummy variables. Otherwise, the fixed effect specification is normally faster to estimate and has cleaner output.

FE Exercise 3: Prescriptions

While many drugs treat one disease, some drugs receive indications (approved uses) for many diseases over their lifecycle. However, once drugs are approved for one disease, doctors can prescribe them to treat essentially any disease.

Suppose you are interested in **the effect in terms of percent change of new indication approvals on the number of fills of prescriptions.** You observe Indication_{idt}, a dummy for whether drug i was approved for disease d by year t, and $\log(\text{Fills}_{idt})$, which reports the number of fills of that drug for that disease in that year.

Propose a model to estimate. What are the advantages and drawbacks of that model?

FE Exercise 3: Prescriptions

Drug fixed effects and year fixed effects.

$$\log(\text{Fills}_{idt}) = \beta_0 + \beta_1 \text{Indication}_{idt} + \alpha_i + \lambda_t + u_{idt}$$

- Controls for time-invariant differences in log fills across drugs and shocks that impact all drugs.
- Does not control for trends in drug- or disease-specific fills. Potentially biased if drugs/diseases that are approved for new uses are trending upwards (or downwards) in use already.
- Effect is estimated using within-drug variation, so it reflects an effect only on the drugs that ever receive an approval.

FE Exercise 3: Prescriptions

Drug-year fixed effects and drug-disease fixed effects.

$$\log(\text{Fills}_{idt}) = \beta_0 + \beta_1 \text{Indication}_{idt} + \alpha_{it} + \gamma_{id} + u_{idt}$$

- Controls for drug-and-time-varying differences in log fills that are fixed across each disease treated by a drug as well as time-invariant differences in fills of drugs for particular diseases.
- Does not control for disease-specific trends, which would still be an issue if diseases that receive approvals are trending upwards in fills.
- Effect still estimated using only within-drug variation, but controlling for drug-specific trends throws away additional within-drug variation, making estimates less precise. May be "asking too much of the data" and not relying enough on reasonable model assumptions.