Topic 6: Fixed Effects, Difference-in-differences, and Factor Models

ECON 5783 — University of Arkansas

Prof. Kyle Butts

November 2024

Estimating the effect of AP classes

Say you want to estimate the effect that taking AP classes in high-school D_i has on the probability of completing a college degree y_i

One concern we might have is that schools that offer AP classes might differ from those that do not.

Estimating the effect of AP classes

Say you want to estimate the effect that taking AP classes in high-school D_i has on the probability of completing a college degree y_i

One concern we might have is that schools that offer AP classes might differ from those that do not.

For example, the average teacher quality Z might be higher in schools that offer AP classes

- Taking AP classes is confounded with attending different quality schools

Omitted Variable Bias

Let $s(i)$ denote the school that student i attended and assume we see multiple students from each school

Omitted Variable Bias

Let $s(i)$ denote the school that student i attended and assume we see multiple students from each school

Say true causal model determining the probability of completing a college degree is given by

$$y_i = D_i \tau + Z_{s(i)} \gamma + u_i$$

- $Z_{s(i)}$ is the quality of the teachers at student i's school
- Assume $\mathbb{E}\left[u_i \mid D_i, Z_{s(i)}\right] = 0$, i.e. there are no other variables correlated with D_i that impact y_i (for the sake of illustration)

Omitted Variable Bias

From our omitted variables bias formula, regressing y_i on D_i (but not $Z_{s(i)}$) would yield

$$\tau_{\text{OLS}} = \tau + \gamma \frac{\text{Cov}\left(D_i, Z_{s(i)}\right)}{\text{Var}\left(D_i\right)}$$

The estimated effect of taking AP classes is biased (up) by the fact that students who take AP classes typically have higher-quality teachers

School indicator variables

The simplest solution would be to measure $Z_{s(i)}$ for each student and control for it using the Selection on Observables tools

- But this variable might not be in our data or might be hard to measure accurately

School indicator variables

The simplest solution would be to measure $Z_{s(i)}$ for each student and control for it using the Selection on Observables tools

- But this variable might not be in our data or might be hard to measure accurately

Instead, consider running the regression of y_i on D_i and a set of indicators for each school $\mathbb{1}[s(i) = k]$ for $k = 1, \ldots, K$:

$$y_i = D_i \tau + \sum_{k=1}^{K} \mathbb{1}[s(i) = k] \alpha_k + u_i$$

- The schools being labeled $1, \ldots, K$ only makes notation easier

School indicator variables

$$y_i = D_i \tau + \sum_{k=1}^{K} \mathbb{1}[s(i) = k]\alpha_k + u_i$$

The set of school indicator variables are often referred to as school 'fixed effects'

- This sum is a bit of a pain to write, so people will often short-hand this model to $y_i = D_i \tau + \alpha_{s(i)} + u_i$ or even just α_s leaving the $s(i)$ implicit.

School indicator variables

To show you what this method does, consider the infeasible regression of y_i on D_i, $Z_{s(i)}$, and indicators for each school

$$y_i = D_i \tau + Z_{s(i)} \gamma + \sum_{k=1}^{K} \mathbb{1}[s(i) = k] \alpha_k + u_i$$

We know from the Frisch-Waugh-Lovell theorem that we can 'residualize' off the school indicator variables and that leaves:

$$\tilde{y}_i = \tilde{D}_i \tau + \tilde{Z}_{s(i)} \gamma + v_i$$

Residualizing $Z_{s(i)}$ on school indicators

What does the residuals of $Z_{s(i)}$ look like? Well it's the regression of

$$Z_{s(i)} = \sum_{k=1}^{K} \mathbb{1}[s(i) = k]\alpha_k + e_i$$

Residualizing $Z_{s(i)}$ on school indicators

What does the residuals of $Z_{s(i)}$ look like? Well it's the regression of

$$Z_{s(i)} = \sum_{k=1}^{K} \mathbb{1}[s(i) = k]\alpha_k + e_i$$

The values of $Z_{s(i)}$ only varies at the school level, so this model will fit the data *perfectly* (set $\alpha_k = Z_k$)

- Consequently, $\tilde{Z}_{s(i)} = 0$ for all observations!

School indicator variables

Therefore, we have that

$$\begin{aligned}\tilde{y}_i &= \tilde{D}_i \tau + \underbrace{\tilde{Z}_{s(i)}}_{=0} \gamma + v_i \\ &= \tilde{D}_i \tau + v_i\end{aligned}$$

So the infeasible regression is, in some sense, actually *feasible*

- We only need to residualize y_i and D_i by the school fixed-effects and the impact of $Z_{s(i)}$ is removed all together

‘Fixed Effects’ and controlling for *unobservables*

This logic extends almost immediately to the case where you have *many* factors that vary at the school-level. Say you have $Z_{1,s(i)}$, $Z_{2,s(i)}$, and $Z_{3,s(i)}$

The FWL theorem logic before is the same

- Residualizing each $Z_{\ell,s(i)}$ separately on school indicator variables will remove them all!

‘Fixed Effects’ and controlling for *unobservables*

This logic extends almost immediately to the case where you have *many* factors that vary at the school-level. Say you have $Z_{1,s(i)}$, $Z_{2,s(i)}$, and $Z_{3,s(i)}$

The FWL theorem logic before is the same

- Residualizing each $Z_{\ell,s(i)}$ separately on school indicator variables will remove them all!

The school fixed-effects absorb *all* the school-level confounders

- The identifying assumption is therefore that there is no *student-level* confounders

‘Fixed Effects’ and controlling for *unobservables*

This is why ‘fixed effects’ are viewed powerfuly in economics and are so commonly used

- They control for many different confounders (that vary at the level of the fixed-effect)
- E.g. any omitted-variables bias story that occurs at the school level is removed by school fixed-effects

What ‘variation’ remains?

So, clearly fixed effects are a powerful tool; but the question is what ‘variation’ remains in $\tilde{D}_i$?

- I.e. what remains after regressing D_i on the set of school indicators

Our regression is $D_i = \sum_{k=1}^{K} \mathbb{1}[s(i) = k]\alpha_k + e_i$

- Predict whether you took an AP class based on the school you go to

What variation remains?

$$D_i = \sum_{k=1}^{K} \mathbb{1}[s(i) = k]\alpha_k + e_i$$

Since the set of school indicators are mutually exclusive and exhaustive, we know that $\hat{\alpha}_k$ estimates the proportion of kids in school k that take AP classes:

- $\hat{\alpha}_k = \mathbb{E}[D_i \mid s(i) = k]$

What variation remains?

$$D_i = \sum_{k=1}^{K} \mathbb{1}[s(i) = k]\alpha_k + e_i$$

Since the set of school indicators are mutually exclusive and exhaustive, we know that $\hat{\alpha}_k$ estimates the proportion of kids in school k that take AP classes:

- $\hat{\alpha}_k = \mathbb{E}[D_i \mid s(i) = k]$

Then the variable $\tilde{D}_i$ equals $D_i - \hat{\alpha}_{s(i)}$

- This takes one of two values within a school: $1 - \hat{\alpha}_{s(i)}$ or $-\hat{\alpha}_{s(i)}$

The intuition is that we are still comparing people with larger or smaller values of D_i, but we are removing bad variation from y_i

Within-school variation

Say you have a relatively small sample of students per school

- In some schools, you either have *everyone* or *no one* that took AP classes

Within-school variation

Say you have a relatively small sample of students per school

- In some schools, you either have *everyone* or *no one* that took AP classes

For these schools, $\hat{\alpha}_{s(i)}$ perfectly fits the data within the school, so $\tilde{D}_i = 0$ for all students in the school

Within-school variation

Say you have a relatively small sample of students per school

- In some schools, you either have *everyone* or *no one* that took AP classes

For these schools, $\hat{\alpha}_{s(i)}$ perfectly fits the data within the school, so $\tilde{D}_i = 0$ for all students in the school

$\implies$ Schools' with no variation in D_i do not contribute to the estimate

Example of Fixed Effects usage

This semester, Sarah Cordes from Temple University presented her work estimating the returns to high-quality schooling on student outcomes among low-income families in New York City

She uses the fact that families are placed randomly within public-housing complexes and are therefore sent to different NYC public schools (of varying quality)

Example of Fixed Effects usage

Cordes et. al. ran a regression of educational gains (y_i) on school quality (q_i) and a set of indicator variables for the public-housing complex a family is assigned to:

$$y_i = q_i \tau + \alpha_c + u_i$$

- α_c denotes complex fixed-effects

Example of Fixed Effects usage

Cordes et. al. ran a regression of educational gains (y_i) on school quality (q_i) and a set of indicator variables for the public-housing complex a family is assigned to:

$$y_i = q_i \tau + \alpha_c + u_i$$

- α_c denotes complex fixed-effects

$\tilde{q}_i = q_i - \mathbb{E}[q_i \mid c = c_i]$ is the within-complex variation in school-quality

- Some families are located near better schools than others within the complex

“Within” estimator

The fixed-effect estimator is sometimes referred to as the ‘within’-estimator.

For example, here is how an economist typically talks about the previous example

- ‘Our estimator compares two kids *within the same public-housing complex*, one assigned to a better than expected quality and the other to a lower than expected quality school’

“Within” estimator

The fixed-effect estimator is sometimes referred to as the ‘within’-estimator.

For example, here is how an economist typically talks about the previous example

- ‘Our estimator compares two kids *within the same public-housing complex*, one assigned to a better than expected quality and the other to a lower than expected quality school’

Note this feels very similar to matching; look within people with the same X_i and argue that treatment is randomly assigned within that group

“Within” estimator

‘Our estimator compares two kids within the same public-housing complex, one assigned to a better than expected quality and the other to a lower than expected quality school’

I do not love this language since it implies (to me) that you are running separate regressions for each public-housing complex (you are not!)

“Within” estimator

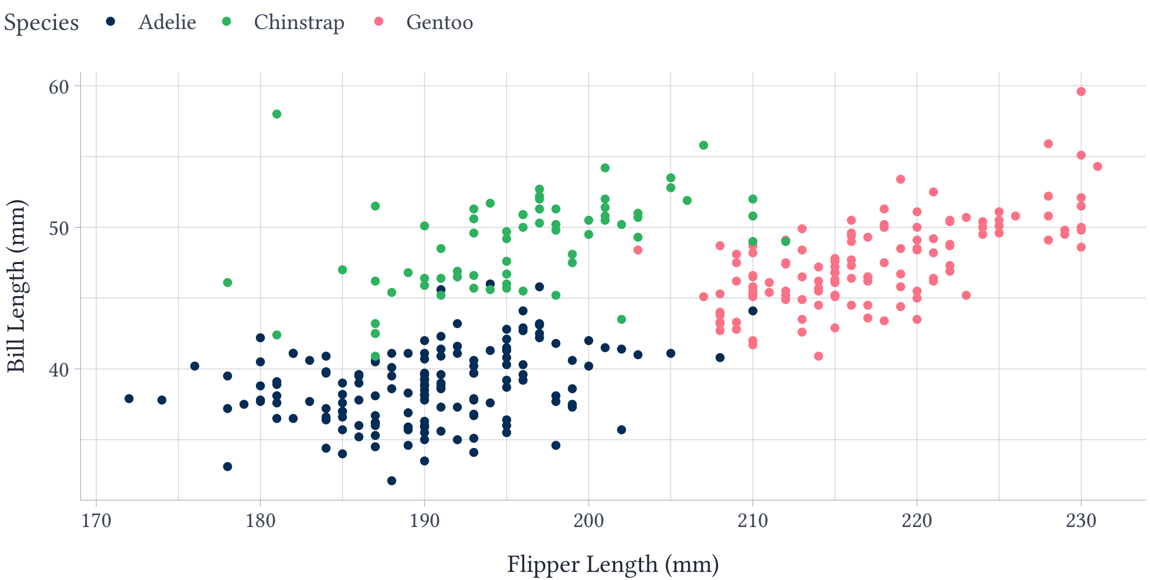
What fixed effects are actually doing:

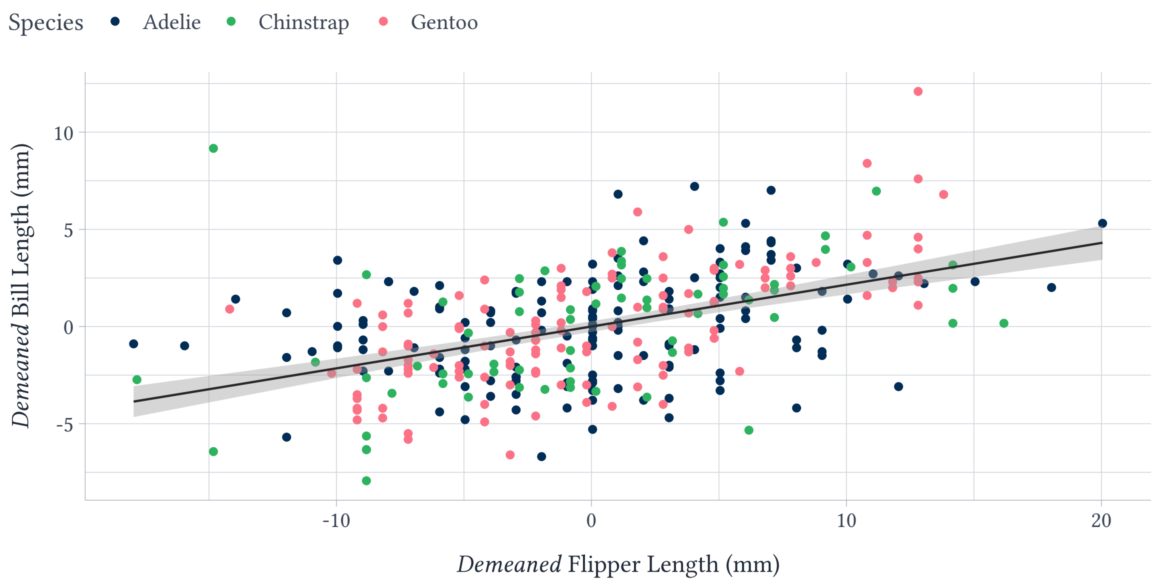
- Within each complex, take deviations between observed q_i and the complex’s average q_i: $\tilde{q}_i$
- Across complexes, regress $\tilde{y}_i$ on $\tilde{q}_i$
  - $\rightarrow$ OLS pools across complexes, putting more weight on public-housing complexes with more variation in $\tilde{q}_i$

Penguins Example

I'll show you an example to try and help make this clearer. I have a dataset with penguin's flipper length and their bill length

- I will regress bill length on flipper length while controlling for species fixed effects (there are 3 different species in this dataset)


Species
Adelie
Chinstrap
Gentoo
60
50
40
Bill Length (mm)
170
180
190
200
210
220
230
Flipper Length (mm)


Species
Adelie
Chinstrap
Gentoo
10
5
0
-5
-10
0
10
20
Demeaned Bill Length (mm)
Demeaned Flipper Length (mm)

“Within” estimator

Am I being a bit pedantic? Kind of...

“Within” estimator

Am I being a bit pedantic? Kind of...; but it can sometimes matter

Remember in our selection on observables section we discussed the weird weighting of units’ treatment effects that can occur with a linear regression:

- In our case, we put more weight on appartment complexes that have more variation in $\tilde{q}_i$ (by least-squares algebra)

“Within” estimator

Am I being a bit pedantic? Kind of...; but it can sometimes matter

Remember in our selection on observables section we discussed the weird weighting of units’ treatment effects that can occur with a linear regression:

- In our case, we put more weight on appartment complexes that have more variation in $\tilde{q}_i$ (by least-squares algebra)
  - $\rightarrow$ In this case, treatment effects probably don’t change much across apartment complex (minimal heterogeneity), so this weighting is probably fine

This will turn out to matter when we discuss difference-in-differences!

Estimating fixed effects in `R`

One final note, to estimate fixed effects in `R`, we will continue to use the `fixest` package:

```
feols(y ~ x | school, data = df, vcov = "hc1")
```

- Put the fixed effects after `|`
- Could do `i(school)`, but will be much slower and clog up your output with extra coefficients

A note on estimation of fixed-effects

Thinking about estimation of fixed-effects, our set of indicator variables can be written as:

$$\begin{pmatrix} \mathbb{1}[s(i) = 1] & \mathbb{1}[s(i) = 2] & \dots & \mathbb{1}[s(i) = S] \end{pmatrix}$$

- This is an $S \times N$ matrix; in larger datasets, your computer would run out of memory creating this matrix

But this matris is very sparse (mostly 0s) and we know what residualizing does:

- In the case of one fixed-effect, we know we can just demean the variables within i
- In the case of multiple fixed-effects, we do something like multiple demeaning

A note on estimation of fixed-effects

Faster methods are implementedin Stata (the O.G.) by `reghdfe` and in R by `fixest`

- It manually demeans y and $\boldsymbol{X}$ and then runs the much simpler $\tilde{y}$ on $\tilde{\boldsymbol{X}}$

In the case of large datasets with many fixed effects, this can take estimation from many hours to a few seconds. Short: use `fixest/reghdfe`

- I find this stuff really interesting, but ymmv

Fixed Effects in Panel Data

Panel Data

Now, we are at the point in the course where we will consider panel data

- We observe a set of individuals $i \in \{1, \dots, N\}$ over a set of time periods $t \in \{1, \dots, T\}$

A balanced panel observers each individual in all T time periods ($N \times T$ total)

- Otherwise, we call this an unbalanced panel

Advantages of Panel Data

In the absence of a clear ‘quasi-experimental’ method in cross-section, people often turn to panel data

Panel data helps us estimate effects by allowing us to remove some key sources of confounding

- Observing a person before they enter into treatment might help us better understand their $Y_{it}(0)$

Caffeine and Productivity

Say you observe a panel dataset of workers (i) on different workdays (t). You want to know if your company should provide free coffee for the workers by evaluating the impact of caffeine (d_{it}) on productivity (y_{it})

The first problem is that worker's that drink coffee probably look different than those that do not

- People with higher (average) d_{it} might differ in other characteristics

Caffeine and Productivity

Say worker's producivity is determined by

$$y_{it} = p_i + \tau d_{it} + \varepsilon_{it}$$

- Here, p_i is a worker's underlying productivity that is time-invariant
- τ is the true treatment effect
- ε_{it} are shocks to productivity on a given day

Unit “Fixed Effects”

$$y_{it} = p_i + \tau d_{it} + \varepsilon_{it}$$

The p_i term is our ‘fixed effect’

- A person-specific effect that is *fixed* over days t
- When we estimate this model, this is just a set of N indicator variables (e.g. a indicator for being person 1, an indicator for being person 2, ...)

Unit “Fixed Effects”

After residualizing out the estimated fixed effects

$$\tilde{y}_{it} = \tau \tilde{d}_{it} + \tilde{\varepsilon}_{it}$$

From least-squares mechanics, we have $\tilde{d}_{it} = d_{it} - \bar{d}_i$, where $\bar{d}_i$ is the (sample) average caffeine intake for a person

- $\tilde{d}_{it}$ represents the average caffeine intake relative to the worker’s average intake

Identifying assumption

$$(y_{it} - \bar{y}_i) = \tau \left(d_{it} - \bar{d}_i\right) + \tilde{\varepsilon}_{it}$$

Our regression compares workers on days where they have above (their) average caffeine intake to days where they have below (their) average

- Our ‘ideal experiment’ is that workers *randomly* have more or less caffeine than their average

Identifying assumption

$$(y_{it} - \bar{y}_i) = \tau \left(d_{it} - \bar{d}_i\right) + \tilde{\varepsilon}_{it}$$

Our 'ideal experiment' is that workers randomly have more or less caffeine than their average

Do we think this is a plausible assumption?

Identifying assumption

$$(y_{it} - \bar{y}_i) = \tau \left(d_{it} - \bar{d}_i\right) + \tilde{\varepsilon}_{it}$$

Our 'ideal experiment' is that workers randomly have more or less caffeine than their average

Do we think this is a plausible assumption?

- Maybe on days when the worker is feeling very tired, they drink an extra cup of coffee
- Them being tired might have a direct effect on their productivity (showing up in $\tilde{\varepsilon}_{it}$)

‘Fixed Effects’ vs. ‘Fixed Characteristics’

A lot of people confuse ‘fixed effects’ with ‘fixed characteristics’

- E.g. your people skills might be a ‘fixed characteristic’
- But over time the labor market returns to people skill might change

So even though your people skills might be fixed, it does not have a ‘fixed *effect*’ on the outcome

‘Fixed Effects’ vs. ‘Fixed Characteristics’

A lot of people confuse ‘fixed effects’ with ‘fixed characteristics’

- E.g. your people skills might be a ‘fixed characteristic’
- But over time the labor market returns to people skill might change

So even though your people skills might be fixed, it does not have a ‘fixed *effect*’ on the outcome

This can cause a bias in our treatment effect

- E.g. If people with high people skills select into treatment when the returns to people skills is going up, that will contaminate treatment effect