Homework for Chapter 16: Fixed Effects

*How Does It Work?*

1. You observe the number of vacations taken by Zac and Skylar in 2012, 2013, and 2014. In those years, Zac took 3, 7, and 5 vacations, respectively. Skylar took 2, 6, and 10.
   1. Isolate the numbers that represent the variation *between* Zac and Skylar in their vacation-taking.

|Mean(Zac)-Mean(Skylar)|=|(3+7+5)/3-(2+6+10)/3|=1

* 1. Isolate the variation *within* Zac and within Skylar in their vacation-taking.

Within(Zac): -2, 2, 0

Within(Skylar): -4, 0, 4

* 1. (Difficult!) We perform a fixed effects analysis of the effect of vacations on happiness. A vacation increases Zac’s happiness by 1 “happiness point,” but it increases Skylar’s happiness by 2 “happiness points.” Will our fixed effects estimate likely give us an answer closer to 1, closer to 2, or exactly 1.5?

Within variation

The fixed effects focus on within variation which makes the treatment effect estimate focuses a lot more heavily on people with a lot of within variation, in this case, that is Skylar. Therefore, the fixed effects estimate will likely give us an answer closer to 2. (15+36)/(15+18) ≈ 1.545

1. You are interested in the effect of cultural events on the levels of trust in a city. Perhaps big events like concerts bring people together and they can trust each other more. You plan to look at the relationship between trust and number of events in a given year, with fixed effects for city. Draw a causal diagram for this research question with at least four back door paths. Which paths will be closed by fixed effects, and which will remain open?

Local Culture

Trust

Local Economy

Number of Cultural Events

Religion of Population

Public Security

Ideally, the four back door paths I list can all be closed by fixed effects for city, among which the local religion is rather constant, and the local culture, economy, and public security are nearly constant and therefore can be nearly controlled for, depending on if any dramatic event happened during the time period and brought any change in the variables.

1. Classify each of the following forms of variation as “between variation”, “within variation”, or a combination of both.
   1. (Individual = person) How a child’s height changes as they age. Within variation
   2. (Individual = person) In a data set tracking many people over many years, the variation in the number of children a person has in a given year. Within variation (can figure out a combination of both between variation and within variation using this data set if we want)
   3. (Individual = city) Overall, Paris, France has more restaurants than Paris, Texas.

Between variation

* 1. (Individual = genre) The average pop music album sells more copies than the average jazz album

Between variation

* 1. (Individual = genre) Miles Davis’ Kind of Blue sold very well for a jazz album.

Within variation. This person’s jazz album (Miles Davis’ Kind of Blue) sold better than other jazz albums.

\* Also can be a combination of both between variation and within variation: The "for a jazz album" implies that we are comparing jazz as a genre to other genres in terms of sales (probably jazz albums don’t usually sell well), so between variation.

* 1. (Individual = genre) Michael Jackson’s Thriller, a pop album, sold many more copies than Kind of Blue, a jazz album.

Between variation

1. Why does the process of taking each observation relative to its individual-level mean have the effect of “controlling for individual”?

If we subtract the individual-level mean from the individual’s observations, we’re left with the way in which the individual varies from time period to time period relative to their own average. This is the within variation, looking at how things vary within individual (whether “individual” in our context means “person,” “company,” “school,” or “country,” and so on). By isolating the within-variation of the individual, we get rid of all the variation between individuals (all the variation in the data explained by individual), thus controlling for individual, i.e., controlling for all variables, whether they’re observed or not, that are constant over time within the individual.

*How is it Performed?*

1. You are interested in the effect of cultural events on the levels of trust in a city. You run a regression of trust levels (on a 0-100 scale) on the number of cultural events with city fixed effects and get a coefficient on cultural events of 3.6. Assume that there are still some back doors open, so do not interpret the result causally. Interpret the 3.6, explaining it in an English sentence.

For a given city, the city’s number of cultural events and trust level are positively associated. When the number of cultural events is one unit higher than it typically is for that city, we'd expect trust levels to be 3.6 points higher than they typically are for that city.

1. You are interested in the effect of cultural events on the levels of trust in a city. You run a regression of trust levels (on a 0-100 scale) on the number of cultural events with city and year fixed effects and get a coefficient on cultural events of 2.4. Assume that there are still some back doors open, so do not interpret the result causally. Interpret the 2.4, explaining it in an English sentence.

For a given city in a given year, there is a positive association between the city’s number of cultural events and trust level. When the number of cultural events is one unit higher than it typically is for that city in that year, then we’d expect levels of trust to be 2.4 units higher than it typically is for that city in that year.

1. Two-way fixed effects with terms for both individual and time are often referred to as “controlling for individual and time effects”. Why might a researcher want to do this rather than just taking individual fixed effects and adding a linear/polynomial/etc. term for time?

If we want to look at variation within individual as well as within a time period, the variation that’s left as being variation relative to what we’d expect given that individual, and given that year. This is not the same as given that individual that time period, as we only observe that individual in that time period once. There’s no variation. Each of the “relative to”s is one at a time. Individual fixed effects affect the year fixed effects and vice versa. By simply adding a linear/polynomial/etc. term for time, we do not achieve the same effect.

1. Which of the following explains why random effects is likely to do a better job of estimating the individual-level effects than fixed effects, if its assumptions hold? c
   1. Because it makes the assumption that the individual effects are unrelated to the other predictors, which breaks that back door and thus reduces bias.
   2. Because random effects allow some amount of between variation into the model, and ~~some of the real individual effect is that between variation~~.
   3. Because it uses the information from the entire data set to estimate each individual effect, rather than relying on only a few observations per individual.
   4. It won’t. Enforcing Durbin-Wu-Hausman makes both methods produce the same estimates anyway.

Coding (which includes any how-the-pros-do-it questions)