**Samuel Snelson**

**SOC 723 Spring 2023**

**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.

Chart, line chart

Description automatically generated

* 1. Isolate the numbers that represent the variation *between* Zac and Skylar in their vacation-taking.
     1. Between-variation is given by the difference in the average number of vacations taken by each person.
     2. Zac has taken 6 vacations on average and Skylar has taken 6 on average. Thus the between-variability is 1 vacation. Skylar’s higher number of vacations in 2014 outruns Zac’s small leads in 2013 and 2014.
  2. Isolate the variation *within* Zac and within Skylar in their vacation-taking.
     1. Zac has an average of 7 and has a within-variability of +2 and -2 (sd = 2)
     2. Skylar as an average of 6 and has a within-variability of +4 and -4 (sd = 4)
  3. (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?
     1. Because Skylar has more within-variability, I am leaning towards the estimate of the effect of vacationing on happiness conditioning on the individual-level fixed effect being be closer to 2 (potentially 1.5 but I’m not confident about exactly).

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?

Fig. 1: DAG for Cultural Events and Trust Across Cities

Diagram

Description automatically generated

* 1. Here are some example confounders for the relationship between the number of cultural events per year and the level of trust in a city. Perceived crime and safety among city residents, infrastructure available to hold events (and interactions), as well as the demographic composition across time in the city are likely to influence both the number of events and level of trust in a city.
  2. Because a city fixed effect concerns time-constant aspects about a city, confounders which are unlikely to change during the period of measurement would be addressed. I am not enthusiastically confident about the constancy of any of these variables in general, but over a few years, this is unlikely to change unless perhaps there is a shift in leading political parties (accelerated efforts to develop infrastructure, possible heightened fears of crime and safety in light of increased attention on immigration).

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.
      1. Within
   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.
      1. Both
   3. (Individual = city) Overall, Paris, France has more restaurants than Paris, Texas.
      1. Between
   4. (Individual = genre) The average pop music album sells more copies than the average jazz album
      1. If we are only looking at the difference in estimated average sales by genre, then just between. If we are comparing the distributions, then both.
   5. (Individual = genre) Miles Davis’ Kind of Blue sold very well for a jazz album.
      1. Within
   6. (Individual = genre) Michael Jackson’s Thriller, a pop album, sold many more copies than Kind of Blue, a jazz album.
      1. Between
2. Why does the process of taking each observation relative to its individual-level mean have the effect of “controlling for individual”?
   1. Similar to how matching controls for the variable that is matched on (when covariates are sufficiently balanced), making comparisons between cases within strata of a variable eliminates the influence of that variable because there isn’t any difference between the comparisons other than the predictors (and, for matching, unobserved confounders, and, for fixed effects, time-variant confounders). There are other considerations to be made as to the functional form, but this is the main idea.

*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.
   1. Adjusting for city fixed effects, for each additional cultural event in a city, the trust level is expected to be about 3.6 points higher relative to the city’s average on a 100 point scale.
2. 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.
   1. Adjusting for city and year fixed effects, for each additional cultural event in a city, the trust level is expected to be about 2.4 points higher relative to the city’s average in a given year.
3. 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?
   1. Fixed effects are a tool for addressing unobserved effect heterogeneity, so this may itself be a reason to try to improve the precision of estimation.
   2. As I understand, using an individual and year fixed effect approach gives some more flexibility to the time-constancy requirement where time-variance is allowed in confounders between years.
4. 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?
   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 allows 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.