What are we trying to do? What is the problem? This doesn’t have an answer

* Causal inference: we are controlling for time-invariant omitted variables between groups (individuals, states, countries)
* Prediction: Handle correlated errors within groups, remove latent differences between units who are sampled randomly from a population
* Where did our data come from?

RCTs vs. observational studies

What do we call a “fixed effect” vs. a “random effect”?

* Random effect = higher-level entity (person, state, country) treated as a distribution or random variable (random unknown)
* Fixed effect = higher-level entity treated as a dummy variable (fixed unknown)

This conversation often starts to blur to line between frequentist and Bayesian statistics, which is itself a whole can of worms

In my opinion, it’s most helpful to think of each *term* in your model this way, as this is how mixed-effects frameworks think about it. Some fields (economics) tend to characterize an *entire model* as a “fixed effects model” or a “random effects model” based on the entire specification, assumed relationships between predictors/residuals, etc.

But this gets tricky because there are ways you can use REs to get the same answer as FEs (e.g., Bell & Jones (2015, 2019): include means of time-varying factors and a random intercept).

* Within-between (WB): Mundlak 1978

REs don’t *necessarily* remove heterogeneity bias arising from unit-level omitted variables – but they can, depending on how your other variables are transformed, what else you include in the model, etc. If you throw a fixed effect into a model, you know you’re controlling for time-invariant omitted variables. That’s not necessarily true for REs.

FEs reflect a methodological framework that is always in reference to the OLS model

* REs can easily be included in generalized linear models (logistic, Poisson, multinomial, etc.)
* They can be as complicated as you want, given your research question and data-generating process
  + When T>2, or temporal autocorrelation; random-walk, autoregressive
  + Spatial autocorrelation; BYM model, GMRF model

Are REs better? No. Key questions for decision-making:

* Is the within-unit variation substantively important to your research question, or are you just trying to get rid of it?
* Just try both: Hausman test
* Journal/field preference

Hi all - following up on a couple items from the TA session.

For anyone who missed it, we reviewed FEs and REs conceptually, including varying definitions and uses across fields. I mentioned the attached paper, which I find to be an extremely helpful review of how different fields define and apply these methods - as well as some practical considerations depending on how many units you have (N) and how many observations you have within units (T). Again, this is not necessary and the conversation of whether FEs or REs are "better" can get pretty esoteric depending on use case, model specification, and data. I think it's most important to start from your research question and thinking through how panel data can help you isolate that question. I'm happy to talk with anyone separately about their specific case as you start to think about class projects.

We also reviewed using the R package ggplot to visualize panel data, using the dataset that will be used in the homework. I've posted the code sharing on my screen as well as a recording in a separate Module.

Nick