Fixed, Random, and Mixed Effects

John Poe, University of Kentucky

Overview

- The semantics of fixed, random, and mixed effects models across fields
- Multilevel data structures
- How you fix the problems
- Building a model around your hypotheses vs your data

Semantics

But the terms “fixed” and “random” mean very different but very specific things to different people

Semantics

The definitions tend to proliferate as people seem to misuse the terms and make up their own definitions from time to time

Semantics

Fixed effects: a type of *model* using only within group variability to estimate model parameters

Semantics

Fixed effects: *variables* that do not vary randomly across groups

Semantics

Fixed effects: *coefficients* on within group varying variables

Semantics

Fixed effects: ***dummy variables*** **used to remove between group variability**

Semantics

Random effects: a latent *variable* made up of the expected values of Y based on group membership

Semantics

Random effects: any *variable* that is allowed to vary across groups within a model

Semantics

Random effects: *the variance* around the model intercept when that intercept is allowed to vary across groups

Semantics

Random effects: ***the variance*** **around any variable that is that is allowed to vary across groups within a model**

Semantics

Random effects: a *class of models* where you allow some parameters to vary across groups (random intercept models, random slope models)

Semantics

Random effects: ***a type of model*** **that causes endogeneity and is basically evil**

Semantics

Mixed effects: ***a type of model*** **that has both random effects and fixed effects**

Semantics

So if someone says fixed & random effects they mean:

- a variable
- a coefficient
- the variance on a coefficient
- multiple variables
- multiple coefficients
- multiple variances around multiple coefficients
- a specific model
- or an entire class of models

Semantics

For our purposes a fixed or random effect will refer to an estimated parameter in a model like an intercept, beta coefficient, or random intercept

Fixed or Random Effects models are very specific types of models that do not include all models with fixed or random parameters in them

Semantics

Pretty straightforward, right?

Data Structures

Most data within the social and behavioral sciences are clustered

Data Structures

Countries/states/counties/firms/people over time

People under the same government

People who interact with one another

People exposed to the same kind of stimulus

People with similar lived experiences

Data Structures

Most of this we can handle easily enough with independent variables

Data Structures

Basic statistical models assume that, conditional on covariates, observations are independent

Data Structures

When your independent variables don’t explain all of the correlation across individuals you have a non-iid data structure

Data Structures

Multilevel or mixed effects data structures are one of the most common ways to think about this

Data Structures

Why should you care *at all* about this?

Data Structures

There are three very distinct problems that can happen

Data Structures

1) **Omitted variable bias messing with the standard errors**
2) **Omitted variable bias confounding X at different levels of analysis because you only have it measured at one level**
3) **Omitted variable bias from the fact that the effect of X is inconsistent across/interactive with groups because you don't have interactions**

Both 2 & 3 are just called endogeneity bias by economists

Data Structures

Your standard errors are probably wrong

Data Structures

Your coefficients are probably wrong

Data Structures

You probably don’t have the right variables in your model

Data Structures

You probably aren’t even testing your hypotheses

Questions?

How you fix it

You have five basic options of equations with various bells and whistles on them in the literature(s)

How you fix it

Different solutions exist depending on which problem(s) you have

How you fix it

Standard errors are easy to fix

How you fix it

- **Hubert-White Cluster Robust Standard Errors**
- **Cluster Bootstrapped/Jackknifed Standard Errors**
- **Including a random effect in the model**

How you fix it

Problems 2 and 3 are more complicated

How you fix it

Multilevel data structures mean that you have variability within groups and between groups and each type of variation needs to be modeled directly

How you fix it

You deal with this by decomposing the within group variability and the between group variability into either multiple error terms, sets of related independent variables, or both

How you fix it

$$Y_i = \alpha + \beta(X_i) + \varepsilon$$

A Classical Linear Regression Model

How you fix it

The fixed effects within the model

$$Y_i = \alpha + \beta(X_i) + \varepsilon$$

A Classical Linear Regression Model

How you fix it

$$Y_{ij} = \alpha + \beta(X_i) + \beta(J_1) + \beta(J_2) + \cdots \beta(J_{g-1}) + \varepsilon$$

Economists:	A Fixed Effects Model
Statisticians:	Very Inefficient
Psychologists:	o_O

How you fix it

The fixed effects within the model

$$Y_{ij} = \alpha + \beta(X_i) + \beta(J_1) + \beta(J_2) + \cdots \beta(J_{g-1}) + \varepsilon$$

Economists:	A Fixed Effects Model
Statisticians:	Very Inefficient
Psychologists:	o_O

How you fix it

$$\boldsymbol{Y_{ij} = \alpha + \beta(X_{ij}) + \mu + \varepsilon}$$

Economists:	A Random Effects Model
Statisticians:	A Random Intercept Model
Psychologists:	A Random Intercept Model

How you fix it

The fixed effects within the model

The random effects within the model

$$Y_{ij} = \alpha + \beta(X_{ij}) + \mu + \varepsilon$$

Economists:	A Random Effects Model
Statisticians:	A Random Intercept Model
Psychologists:	A Random Intercept Model

How you fix it

$$\boldsymbol{Y_{ij} = \alpha + \beta(X_{ij} - \bar{X}_j) + \beta\bar{X}_j + \mu + \varepsilon}$$

Economists:	A Mundlak Device
Psychologists:	Group Mean Centering
Statisticians:	Group Mean Centering

How you fix it

The fixed effects within the model

The random effects within the model

$$Y_{ij} = \alpha + \beta(X_{ij} - \bar{X}_j) + \beta\bar{X}_j + \mu + \varepsilon$$

Economists:	A Mundlak Device
Psychologists:	Group Mean Centering
Statisticians:	Group Mean Centering

How you fix it

$$Y_{ij} = \alpha + \beta(X_{ij} - \bar{X}_{ij}) + \beta(\bar{X}_j - \bar{X}_{ij}) + \mu + \varepsilon$$

Economists:	Witchcraft
Psychologists:	Grand Mean Centering
Statisticians:	Grand Mean Centering

How you fix it

The fixed effects within the model

The random effects within the model

$$Y_{ij} = \alpha + \beta(X_{ij} - \bar{X}_{ij}) + \beta(\bar{X}_j - \bar{X}_{ij}) + \mu + \varepsilon$$

Economists:	Witchcraft
Psychologists:	Grand Mean Centering
Statisticians:	Grand Mean Centering

How you fix it

$$\boldsymbol{Y_{ij} = \alpha + \beta(X_{ij} - \bar{X}_j) + \beta\bar{X}_j + \mu + \mu(X_{ij}) + \varepsilon}$$

Everyone: A Random Coefficients Model
Everyone: A Random Slopes Model

How you fix it

The fixed effects within the model

The random effects within the model

$$Y_{ij} = \alpha + \beta(X_{ij} - \bar{X}_j) + \beta\bar{X}_j + \mu + \mu(X_{ij}) + \varepsilon$$

Everyone: A Random Coefficients Model

Everyone: A Random Slopes Model

How you fix it

Briefly recapping

- **Fixed effects models (dummy variables)**
- **Random effects models (just the RE)**
- **Multilevel Models v1 (A Mundlak set up)**
- **Multilevel Models v2 (grand mean centering)**
- **Random coefficients (interactive multilevel models)**

How you fix it

Briefly recapping some more

- **Decomposing your error term (all but FE model)**
- **Decomposing your independent variables (all but RE model)**
- **Interacting your decomposed variables and decomposed error term (RC model)**

How you fix it

Each of these is providing some type of solution for omitted variable bias by including measures for the grouping structure

How you fix it

Your choice is based on your hypotheses and your data

Use Fixed Effects

Hypothesis Considerations:

1) **You only care about within group variability and not between group variability**
2) **You care about comparing specific groups**

Use Fixed Effects

Your hypotheses are about within person change over time or average within group effects controlling for average group differences

Use Fixed Effects

This is what you are looking for in simple treatment evaluation studies and is why fixed effects estimators are usually taught in the context of causal inference

Use Fixed Effects

Data Considerations:

1) You don’t have many groups
2) You can’t figure out how to specify the right kind of model with random effects
3) You would otherwise need to use a bunch of random coefficients and the model isn’t computationally stable

Use Random Effects

Hypothesis Considerations:

1) **If you have hypotheses about group level variables that are time or group invariant**

A Model from Theory

- Gender and race (in panel data)
- Group characteristics like a country's region or GDP (in cross-sectional data)

Use Random Effects

Data Considerations:

1) **You have no correlation between independent variables and the random effect**
2) **You are running a nonlinear maximum likelihood model and want to be careful**

Use a Mundlak device

Hypothesis Considerations:

1) **You care about understanding**
   - **contextual effects**
   - **group level variables**
   - **within group effects**
   - **cross-level effects**

Use a Mundlak Device

Data Considerations:

1) **You want to do a Fixed Effects Model but you have a (very) nonlinear outcome**
2) **You don't have much within group variability**
3) **You plan to use random coefficients**

Use Random Coefficients

Hypothesis Considerations:

1) **You care about understanding how the effect of an independent variable varies across groups**

Use Random Coefficients

Pretty much anytime you want to understand context effects

Use Random Coefficients

Data Considerations:

1) **The Mundlak device still isn't getting you unbiased within group coefficients**
    a) **And it's either a nonlinear model or you care about group/time invariant variables**

Use Random Coefficients

You want to know how people are different across different contexts

Wall of citations: Books

Panel/Longitudinal

- Hsiao, Cheng. 2014. *Analysis of panel data*. Cambridge university press.
- Baltagi, Badi. 2013. *Econometric analysis of panel data*: Wiley.
- Fitzmaurice, Garrett M, Nan M Laird, and James H Ware. 2012. *Applied longitudinal analysis*. John Wiley & Sons.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second ed.

Multilevel

- Hox, Joop. 2010. *Multilevel analysis: Techniques and applications*: Routledge.
- Snijders, Tom AB. 2011. *Multilevel analysis*: Springer.
- Goldstein, Harvey. 2011. *Multilevel statistical models*. John Wiley & Sons.
- de Leeuw, Jan, and Erik Meijer. 2008. *Handbook of multilevel analysis*: Springer.
- Wu, Lang. 2009. *Mixed effects models for complex data*: CRC Press.
- Hox, Joop, and J Kyle Roberts. 2011. *Handbook of advanced multilevel analysis*: Psychology Press.
- Scott, Marc A, Jeffrey S Simonoff, and Brian D Marx. 2013. *The SAGE handbook of multilevel modeling*: Sage.

Wall of citations: Mundlak/Centering

Mundlak, Yair. 1978. "On the pooling of time series and cross section data." *Econometrica: Journal of the Econometric Society*:69-85.

Hausman, Jerry A, and William E Taylor. 1981. "Panel data and unobservable individual effects." *Econometrica: Journal of the Econometric Society*:1377-98.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second ed. See chapter 10.7.3 for a detailed discussion of the Hausman specification test with reference to the Mundlak device.

Hox, Joop. 2010. *Multilevel analysis: Techniques and applications*: Routledge. See chapter 4.2 on centering and standardizing explanatory variables, 4.4 on group mean centering

Snijders, Tom AB. 2011. *Multilevel analysis*: Springer. See chapter 4.6 on within and between group regressions, 5.2.2 on the specification of random slope models.

Baltagi, Badi. 2013. *Econometric analysis of panel data*: Wiley. See chapter 4.3 on the Hausman Specification Test, 7.4-7.6 on the Hausman-Taylor (Mundlak) estimator that works as an alternative to strict FE/RE approaches.

Bell, Andrew, and Kelvyn Jones. 2015. "Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data." *Political Science Research and Methods* 3 (01):133-53.

Wall of citations: FE or RE

General Choices

- Nerlove, Marc. 2000. An essay on the history of panel data econometrics. Paper read at Proceedings of Ninth International Conference on Panel Data, Geneva, Switzerland.
- Beck, N., and J. N. Katz. 2007. "Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments." *Political Analysis* 15 (2):182-95.
- Clark, Tom S, and Drew A Linzer. 2012. "Should I Use Fixed or Random Effects?" *Unpublished paper*.
- Bartels, Brandon. "Beyond "Fixed versus Random Effects": A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel Data."
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second ed. Chapters 10.7.2-10.7.3 and 11.2-11.3

FE in Nonlinear models

- Lancaster, Tony. 2000. "The incidental parameter problem since 1948." *Journal of Econometrics* 95 (2):391-413.
- Katz, Ethan. 2001. "Bias in conditional and unconditional fixed effects logit estimation." *Political Analysis* 9 (4):379-84.
- Greene, William. 2004. "The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects." *The Econometrics Journal* 7 (1):98-119.
- Beck, Nathaniel. 2015. Estimating grouped data models with a binary dependent variable and fixed effects: What are the issues? Paper read at annual meeting of the Society for Political Methodology, July.

Random Coefficients

- Beck, N., and J. N. Katz. 2007. "Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments." *Political Analysis* 15 (2):182-95.
- Snijders, Tom AB. 2011. *Multilevel analysis*: Springer. See chapter 5
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second ed. Chapter 4.3.3 and 6.4