UNIVERSITY OF SOUTH CAROLINA
Darla Moore School of Business

Panel Data: Methodological Frameworks and Analytic Tools

Paul D. Bliese

Purpose

- Panel data is common in organizational research, and can be approached using a surprising large number of methodological frameworks and analytic tools
  - Discuss methodological frameworks
  - Compare and contrast tools used by macro and micro-oriented researchers

UNIVERSITY OF SOUTH CAROLINA
Darla Moore School of Business

Purpose

Journal of Management
Vol. XX No. X, Month XXXX 1–30
DOI: 10.1177/0149206319868016

Bridging Methodological Divides Between Macro- and Microresearch: Endogeneity and Methods for Panel Data

Paul D. Bliese
Donald J. Schepker
Spenser M. Essman
Robert E. Ployhart
University of South Carolina

Both macro- and micro-oriented researchers frequently use panel data where the outcome of interest is measured repeated times. Panel data support at least five different modeling frameworks (within, between, incremental/emergent, cross-level, and growth). Researchers from macro- and micro-oriented domains tend to differentially use the frameworks and also use different analytic tools and terminology when using the same modeling framework. These differences have the potential to inhibit cross-discipline communication. In this review, we explore how macro- and microresearchers approach panel data with a specific emphasis on the theoretical implications of choosing one framework versus another. We illustrate how fixed-effects and random-effects models differ and how they are similar, and we conduct a thorough review of 142 articles that used panel data in leading management journals in 2017. Ultimately, our review identifies ways that researchers can better employ fixed- and random-effects models, model time as a meaningful predictor or ensure unobserved time heterogeneity is controlled, and align hypotheses to analytic choice. In the end, our goal is to help facilitate communication and theory development between macro- and micro-oriented management researchers.

Keywords: *methodology; endogeneity; panel data; random effects; fixed effects*

Outline

- Purpose
- Panel Data
- Framework 1: Within-Effects
- Framework 2: Between-Effects
  - Inferences and Fallacies
- Framework 3: Incremental or Emergent Models
- Framework 4: Cross-level Interactions
- Framework 5: Growth Models
- Summary
- Conclusion

Panel Data

- Each higher-level entity (ID) provides multiple measures
  - Longitudinal design

- Broad research question:
  - Are perceptions of leadership related to depression?

ID	TIME	DEPRESS	LEAD
1	0	1.8	3
1	1	1.8	3
1	2	3.8	3
38	0	3.4	4
38	1	2.8	3
38	2	2.8	3
51	0	2.2	2
51	1	2	3
51	2	1.6	3

Endogeneity

- With this data, the leadership variable is likely related to the models' error variance (endogeneity)

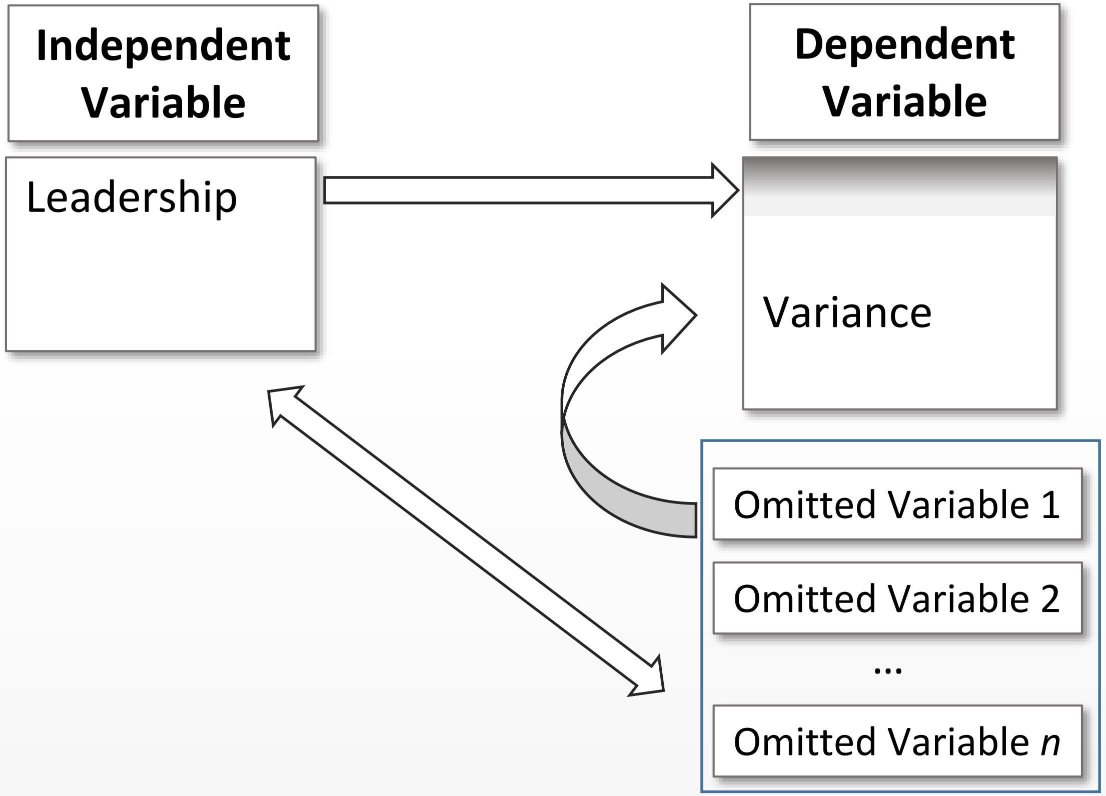
- Micro terms
  - Unmeasured variables
  - Reverse causation
  - Single source variance
  - Common method variance

ID	TIME	DEPRESS	LEAD
1	0	1.8	3
1	1	1.8	3
1	2	3.8	3
38	0	3.4	4
38	1	2.8	3
38	2	2.8	3
51	0	2.2	2
51	1	2	3
51	2	1.6	3

Endogeneity

- Endogeneity occurs when model error is related to the independent variable

- Hard to estimate the true relationship between leadership and depression



Endogeneity

- Without randomization impossible to eliminate all endogeneity

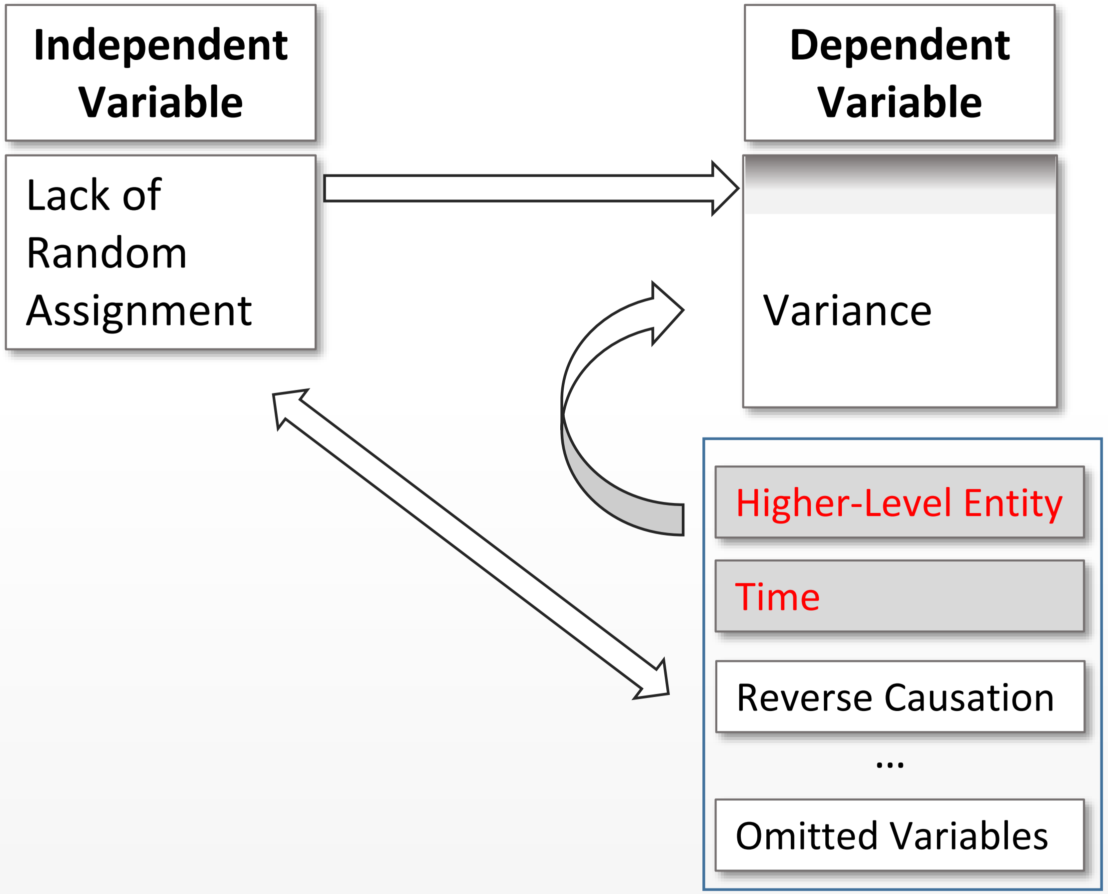
- Possible to eliminate two key sources
  - Unmeasured heterogeneity due to higher-level entity (ID)
  - Unmeasured heterogeneity due to temporal shocks (TIME)

ID	TIME	DEPRESS	LEAD
1	0	1.8	3
1	1	1.8	3
1	2	3.8	3
38	0	3.4	4
38	1	2.8	3
38	2	2.8	3
51	0	2.2	2
51	1	2	3
51	2	1.6	3

Endogeneity

- A properly specified model explaining "within-effects" eliminates two sources of endogeneity

- Our running example



Endogeneity

- Our broad research question:
  - Are perceptions of leadership related to depression?

ID	TIME	DEPRESS	LEAD
1	0	1.8	3
1	1	1.8	3
1	2	3.8	3
38	0	3.4	4
38	1	2.8	3
38	2	2.8	3
51	0	2.2	2
51	1	2	3
51	2	1.6	3

Do Not Do This

- A statistical model everyone agrees is a bad idea
  - Regress Depression on Leadership (OLS)

```
> tmod<-lm(DEPRESS~LEAD,data=CARMA2019)
> round(summary(tmod)$coef,dig=3)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.601      0.065  55.231        0
LEAD          -0.246      0.020 -12.207        0
```

  - To use this model you have to “pretend” each row is independent (which takes a lot of pretending)
    - Each of the **743** respondents provides 3 rows of data and data was collected at 3 different time points

Framework 1: The Within-Effects Model

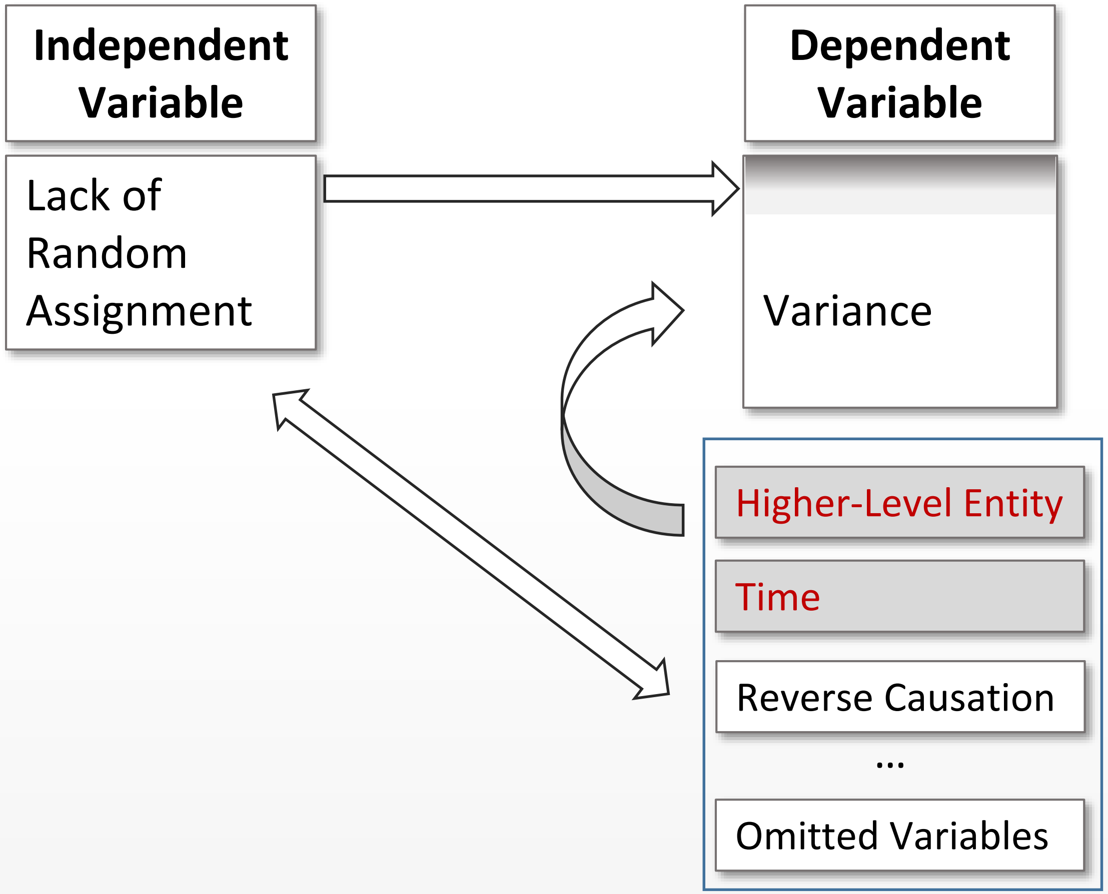
- A fixed-effects model controlling for higher-level entity (ID) and TIME

```
> tmod<-lm(DEPRESS~LEAD+factor(TIME)+factor(ID),data=CARMA2019)
> round(summary(tmod)$coef[1:7,],dig=3)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.081      0.434   7.099    0.000
LEAD            -0.173      0.021  -8.045    0.000
factor(TIME)1   -0.076      0.039  -1.935    0.053
factor(TIME)2   -0.210      0.039  -5.311    0.000
factor(ID)38     0.591      0.605   0.977    0.329
factor(ID)51    -0.591      0.605  -0.977    0.329
factor(ID)55     0.236      0.605   0.390    0.697
AND LOTS MORE…742 dummy codes for ID in total
```

Framework 1: The Within-Effects Model

- Estimate that includes unexplained heterogeneity due to higher-level entity and time
  - Value of **-0.246**

- Estimate removing heterogeneity due to higher-level entity and time
  - Value of **-0.173**



Framework 1: The Within-Effects Model

- The fixed-effects estimate of **-0.173** explains only within-entity variance

- A clear hypothesis will state something like:
  - Within-person changes in perceptions of leadership are related to within-person changes in depression
  - A person's relative perceptions of leadership are related to his or her relative levels of depression

Framework 1: The Within-Effects Model

- One important starting question
  - Is there sufficient within-entity variance?
- In our example, ICC(1) is .448, so 55.2% of variance is within-person

```
> library(nlme)
> tmod<-lme(DEPRESS~1,random=~1|ID,CARMA2019)
> VarCorr(tmod)
ID = pdLogChol(1)
            Variance  StdDev
(Intercept) 0.4683603 0.6843686
Residual    0.5771377 0.7596958
> 0.4683603/(0.4683603+0.5771377)
[1] 0.4479782
```

Framework 1: The Within-Effects Model

- Might refine ICC(1) estimate with an ICC(1) conditional on TIME
  – Minor change from 0.448 to 0.451

```
> tmod<-lme(DEPRESS~factor(TIME),random=~1|ID,CARMA2019)
> VarCorr(tmod)
ID = pdLogChol(1)
            Variance  StdDev
(Intercept) 0.4700021 0.6855670
Residual    0.5722117 0.7564468
> 0.4700021/(0.4700021+0.5722117)
[1] 0.4509651
```

Framework 1: The Within-Effects Model

- The ICC(1) is equivalent to the adjusted r-square in regression

- An OLS model that estimates how much variance is explained by person

```
> tmod<-lm(DEPRESS~factor(ID),CARMA2019)
> summary(tmod)$adj.r.squared
[1] 0.447756
```

- An estimate of how much variance is explained by person and time

```
> tmod<-lm(DEPRESS~factor(ID)+factor(TIME),CARMA2019)
> summary(tmod)$adj.r.squared
[1] 0.4524696
```

Framework 1: The Within-Effects Model

- The within-effects model can be estimated using:
  - Fixed-effects OLS (our example)
  - Random-effects model (not yet illustrated)
- A limitation with using a random-effects model is that it will not default to the “gold-standard” estimate of -0.173
  - Many options to make a random-effects model return the gold-standard

Framework 1: The Within-Effects Model

- A random-effects model with fully crossed random effects for ID and TIME

```
> library(lme4)
> tmod<-lmer(DEPRESS~LEAD+(1|TIME)+(1|ID),CARMA2019)
> round(summary(tmod)$coef,dig=3)
            Estimate Std. Error t value
(Intercept)    3.498      0.091  38.476
LEAD          -0.212      0.019 -11.239
```

Framework 1: The Within-Effects Model

- A random-effects model for ID and fixed-effects for TIME
  - A nice way to integrate fixed and random effects models
  - Panel data usually have relatively few time points so it is practical to treat time as a factor (nominal variable)

```
> library(nlme)
> tmod<-lme(DEPRESS~LEAD+factor(TIME),random=~1|ID,CARMA2019)
> round(summary(tmod)$tTable,dig=3)
                Value Std.Error   DF t-value p-value
(Intercept)     3.608     0.073 1483  49.649    0.00
LEAD           -0.214     0.019 1483 -11.290    0.00
factor(TIME)1 -0.091     0.039 1483  -2.324    0.02
factor(TIME)2 -0.227     0.039 1483  -5.777    0.00
```

Framework 1: The Within-Effects Model

- Hausman test if fixed-effects and random-effects models differ

```
> library(plm)
> wi<-plm(DEPRESS~LEAD+factor(TIME),index=c("ID"),data=CARMA2019,
        model="within")
> #summary(wi)$coef
> re<-plm(DEPRESS~LEAD+factor(TIME),index=c("ID"),data=CARMA2019,
                model="random")
> #summary(re)$coef
> phtest(wi,re)
      Hausman Test
data:  DEPRESS ~ LEAD + factor(TIME)
chisq = 16.332, df = 3, p-value = 0.0009693
alternative hypothesis: one model is inconsistent
```

Framework 1: The Within-Effects Model

- Four ways for a random-effects model to recover the gold-standard
  - Fixed-effect for TIME and add person-means for leadership
  - Fixed-effect for TIME and demean (i.e., group-mean center) leadership
  - Fixed-effect for TIME, demean leadership, add person-means for leadership (Hybrid Approach highlighted by Certo et al. 2017)
  - Fully crossed random effects for TIME and ID and person-means for leadership and TIME-means for leadership in model

Certo, S. T., Withers, M. C., & Semadeni, M. 2017. A tale of two effects: Using longitudinal data to compare within and between-firm effects. *Strategic Management Journal*, 38: 1536-1556.

Data Structure with Means

- Person-Mean added back for leadership (LEAD.PM)

ID	TIME	DEPRESS	LEAD	LEAD.PM
1	0	1.8	3	3.00
1	1	1.8	3	3.00
1	2	3.8	3	3.00
38	0	3.4	4	3.33
38	1	2.8	3	3.33
38	2	2.8	3	3.33
51	0	2.2	2	2.67
51	1	2	3	2.67
51	2	1.6	3	2.67

Framework 1: The Within-Effects Model

- A random-effects model for ID and fixed-effects for TIME
  - Add person means for Leadership (LEAD.PM)

```
> tmod<-lme(DEPRESS~LEAD+LEAD.PM+factor(TIME),
  random=~1|ID,CARMA2019)
> round(summary(tmod)$tTable,dig=3)
                 Value Std.Error   DF t-value p-value
(Intercept)      4.023     0.126 1483  31.857   0.000
LEAD            -0.173     0.021 1483  -8.045   0.000
LEAD.PM         -0.179     0.045  741  -3.995   0.000
factor(TIME)1 -0.076     0.039 1483  -1.935   0.053
factor(TIME)2 -0.210     0.039 1483  -5.311   0.000
```

UNIVERSITY OF SOUTH CAROLINA
Darla Moore School of Business

Data Structure with Means

- Creating a demeaned (i.e., group-mean-centered) leadership variable

ID	TIME	DEPRESS	LEAD	LEAD.PM	LEAD.DE
1	0	1.8	3	3.00	0
1	1	1.8	3	3.00	0
1	2	3.8	3	3.00	0
38	0	3.4	4	3.33	0.67
38	1	2.8	3	3.33	-0.33
38	2	2.8	3	3.33	-0.33
51	0	2.2	2	2.67	-0.67
51	1	2	3	2.67	0.33
51	2	1.6	3	2.67	0.33

Framework 1: The Within-Effects Model

- A random-effects model for ID and fixed-effects for TIME
  - Demeaned leadership (LEAD.DE)

```
> tmod<-lme(DEPRESS~LEAD.DE+factor(TIME),random=~1|ID,CARMA2019)

> round(summary(tmod)$tTable,dig=3)
                Value Std.Error   DF t-value p-value
(Intercept)     2.942     0.038 1483  78.253   0.000
LEAD.DE        -0.173     0.021 1483  -8.045   0.000
factor(TIME)1 -0.076     0.039 1483  -1.935   0.053
factor(TIME)2 -0.210     0.039 1483  -5.311   0.000
```

Data Structure with Means

- A model with person means and demeaned leadership

ID	TIME	DEPRESS	LEAD	LEAD.PM	LEAD.DE
1	0	1.8	3	3.00	0
1	1	1.8	3	3.00	0
1	2	3.8	3	3.00	0
38	0	3.4	4	3.33	0.67
38	1	2.8	3	3.33	-0.33
38	2	2.8	3	3.33	-0.33
51	0	2.2	2	2.67	-0.67
51	1	2	3	2.67	0.33
51	2	1.6	3	2.67	0.33

Framework 1: The Within-Effects Model

- A random-effects model for ID and fixed-effects for TIME
  - Demeaned leadership (LEAD.DE) and person means (LEAD.PM)
  - Certo et al., (2017) Hybrid model

```
> tmod<-lme(DEPRESS~LEAD.DE+LEAD.PM+factor(TIME),random=~1|ID,
  data=CARMA2019)

> round(summary(tmod)$tTable,dig=3)
                 Value Std.Error   DF t-value p-value
(Intercept)      4.023     0.126 1483  31.857   0.000
LEAD.DE         -0.173     0.021 1483  -8.045   0.000
LEAD.PM         -0.352     0.039  741  -8.937   0.000
factor(TIME)1   -0.076     0.039 1483  -1.935   0.053
factor(TIME)2   -0.210     0.039 1483  -5.311   0.000
```

Data Structure with Means

- Data for fully crossed model with person means and TIME means

ID	TIME	DEPRESS	LEAD	LEAD.PM	LEAD.DE	LEAD.TM
1	0	1.8	3	3.00	0	3.33
1	1	1.8	3	3.00	0	2.96
1	2	3.8	3	3.00	0	2.91
38	0	3.4	4	3.33	0.67	3.33
38	1	2.8	3	3.33	-0.33	2.96
38	2	2.8	3	3.33	-0.33	2.91
51	0	2.2	2	2.67	-0.67	3.33
51	1	2	3	2.67	0.33	2.96
51	2	1.6	3	2.67	0.33	2.91

Framework 1: The Within-Effects Model

- A fully crossed random effects for ID and TIME
  - Person means (LEAD.PM) and TIME means (LEAD.TM) as covariates

```
> library(lme4)
> tmod<-lmer(DEPRESS~LEAD+LEAD.PM+LEAD.TM+(1|TIME)+(1|ID),
  data=CARMA2019)
> round(summary(tmod)$coef,dig=3)
            Estimate Std. Error t value
(Intercept)    2.727      0.764   3.568
LEAD          -0.173      0.021  -8.045
LEAD.PM       -0.179      0.045  -3.995
LEAD.TM        0.391      0.246   1.594
```

Framework 2: A Between-Effects Model

- In our example (and in most panel data) there is a substantial amount of between-entity variance

- Broad research question of “Are perceptions of leadership related to depression?” lets us consider between-person relationships
  - Broad question good for this presentation but poor practice in terms of good science

Framework 2: A Between-Effects Model

- A between-effects OLS regression model based on person-means
  - Value of **-0.352** is roughly twice the size of within effect of **-0.173**

```
> AV.DAT<-aggregate(CARMA2019[,c("ID","LEAD","DEPRESS")],
  by=list(CARMA2019$ID),mean)
> nrow(AV.DAT)
[1] 743
> AV.DAT<-AV.DAT[,2:4]
> names(AV.DAT)<-c("ID","LEAD.PM","DEPRESS.PM")

> tmod<-lm(DEPRESS.PM~LEAD.PM,data=AV.DAT)
> round(summary(tmod)$coef,dig=3)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.928      0.124  31.627        0
LEAD.PM       -0.352      0.039  -8.937        0
```

UNIVERSITY OF SOUTH CAROLINA
Darla Moore School of Business

Framework 2: A Between-Effects Model

- In a random-effects model, regress depression at each time point on person-means for leadership

ID	TIME	DEPRESS	LEAD	LEAD.PM
1	0	1.8	3	3.00
1	1	1.8	3	3.00
1	2	3.8	3	3.00
38	0	3.4	4	3.33
38	1	2.8	3	3.33
38	2	2.8	3	3.33

Framework 2: A Between-Effects Model

- Person mean for leadership **does not** explain within person change in depression (Bliese et al., 2018; LoPilato & Vandenberg, 2015)
  - Great way to conduct an analysis of group means in a random-effects model

```
> tmod<-lme(DEPRESS~LEAD.PM,random=~1|ID,CARMA2019)
> round(summary(tmod)$tTable,dig=3)
             Value Std.Error   DF t-value p-value
(Intercept)  3.928     0.124 1486  31.627       0
LEAD.PM     -0.352     0.039  741  -8.937       0
```

Bliese, P. D., Maltarich, M. A., & Hendricks, J. L. (2018). Back to Basics with Mixed-Effects Models: Nine Take-Away Points. *Journal of Business and Psychology, 33*, 1-23.

LoPilato, A.C., & Vandenberg, R.J. (2015). The not so direct cross-level direct effect. In C.E. Lance & R.J. Vandenberg (Eds.), *More Statistical and Methodological Myths and Urban Legends* (pp. 292-310). New York, NY: Routledge.

University of South Carolina
Darla Moore School of Business

Framework 2: A Between-Effects Model

- Use demeaned leadership (LEAD.DE) and person-means of leadership (LEAD.PM) to estimate both within and between simultaneously

ID	TIME	DEPRESS	LEAD	LEAD.PM	LEAD.DE
1	0	1.8	3	3.00	0
1	1	1.8	3	3.00	0
1	2	3.8	3	3.00	0
38	0	3.4	4	3.33	0.67
38	1	2.8	3	3.33	-0.33
38	2	2.8	3	3.33	-0.33
51	0	2.2	2	2.67	-0.67
51	1	2	3	2.67	0.33
51	2	1.6	3	2.67	0.33

Framework 2: A Between-Effects Model

- Demeaned leadership (LEAD.DE) and person-means of leadership (LEAD.PM) estimate both within-effects and between-effects
  – Hybrid Model (Certo et al., 2017)

```
> tmod<-lme(DEPRESS~LEAD.DE+LEAD.PM+factor(TIME),random=~1|ID,
  CARMA2019)
> round(summary(tmod)$tTable,dig=3)
                  Value Std.Error   DF t-value p-value
(Intercept)       4.023     0.126 1483  31.857   0.000
LEAD.DE          -0.173     0.021 1483  -8.045   0.000
LEAD.PM          -0.352     0.039  741  -8.937   0.000
factor(TIME)1 -0.076     0.039 1483  -1.935   0.053
factor(TIME)2 -0.210     0.039 1483  -5.311   0.000
```

Framework 2: A Between-Effects Model

- The between-effects model is only explaining differences among higher-level entities

- A clear hypothesis will state something like this:
  - Average leadership ratings from individuals will be negatively related to their average ratings of depression
  - Individuals with high average ratings of their leaders will tend to have low average ratings of depression

Inferences and Fallacies

- Both within and between estimates are correct
  - To describe how changes in leadership relate to changes in depression interpret the value of -0.173
  - To describe how average levels of depression relate to overall ratings of leadership interpret the value of -0.352

- Ecological fallacy – use the value of -0.352 to make an inference about how changes over time in leadership are related to changes in depression
- Atomistic fallacy – use the value of -0.173 to make an inference about average levels of depression based on average leadership ratings

Framework 3: Incremental or Emergent

- In many cases, the question of whether a person's average ratings of leadership are related to a person's average ratings of depression is not very interesting

- If within-effects are significant, person means will "inherit" within effect even if "persons" are randomly created

Framework 3: Incremental or Emergent

- Create pseudo people
  - Randomly select any [DEPRESS, LEAD.DE] pair at TIME 0, TIME 1, and TIME 2 combine into pseudo person 1
  - Repeat 743 times (without replacement)
- Create pseudo person means
- Estimate between-effect
  - Recall, within-effect is -0.173

ID	TIME	DEPRESS	LEAD.DE
1	0	1.8	0
1	1	1.8	0
1	2	3.8	0
38	0	3.4	0.67
38	1	2.8	-0.33
38	2	2.8	-0.33
51	0	2.2	-0.67
51	1	2	0.33
51	2	1.6	0.33

Framework 3: Incremental or Emergent

- Between-effects from 1000 runs
  - Each run creates 743 pseudo person-means for depression and leadership

```
ran.pan<-function(dat){
  RDAT<-rbind(dat[dat$TIME==0,c("DEPRESS","LEAD.DE")][sample(1:743),],
       dat[dat$TIME==1,c("DEPRESS","LEAD.DE")][sample(1:743),],
       dat[dat$TIME==2,c("DEPRESS","LEAD.DE")][sample(1:743),])
 RDAT$TIME<-rep(0:2,each=743)
 RDAT$ID<-rep(1:743,3)
 TDAT<-aggregate(RDAT[,c("ID","DEPRESS","LEAD.DE")],list(RDAT$ID),mean)
 tmod<-lm(DEPRESS~LEAD.DE,data=TDAT)
 OUT<-summary(tmod)$coef[2,1]
 return(OUT)}

> ROUT<-replicate(1000,ran.pan(CARMA2019))
> mean(ROUT)
[1] -0.1715586
```

Framework 3: Incremental or Emergent

- Recall, one way to recover the "gold-standard" within-effect in a random effects model is to include the person-level mean of the predictor

ID	TIME	DEPRESS	LEAD	LEAD.PM
1	0	1.8	3	3.00
1	1	1.8	3	3.00
1	2	3.8	3	3.00
38	0	3.4	4	3.33
38	1	2.8	3	3.33
38	2	2.8	3	3.33
51	0	2.2	2	2.67
51	1	2	3	2.67
51	2	1.6	3	2.67

UNIVERSITY OF SOUTH CAROLINA
Darla Moore School of Business

Framework 3: Incremental or Emergent

- LEAD.PM value represents incremental or emergent effect
  - The between-effect is significantly different (-0.179 stronger) than the within-effect
  - Total effect is -0.173+-0.179 = -0.352

```
> tmod<-lme(DEPRESS~LEAD+LEAD.PM+factor(TIME),
  random=~1|ID,CARMA2019)
> round(summary(tmod)$tTable,dig=3)
                 Value Std.Error   DF t-value p-value
(Intercept)      4.023     0.126 1483  31.857   0.000
LEAD            -0.173     0.021 1483  -8.045   0.000
LEAD.PM         -0.179     0.045  741  -3.995   0.000
factor(TIME)1 -0.076     0.039 1483  -1.935   0.053
factor(TIME)2 -0.210     0.039 1483  -5.311   0.000
```

Framework 3: Incremental or Emergent

- A clear hypotheses for this model would state something like
  - The relationship at the person-level is significantly stronger that the relationship at the within-person level
- Interesting model
  - Rarely developed on theoretical grounds but common (Bliese et al., in press; Certo et al., 2017)
- Likely different literature streams and different (but complementary) mechanisms would form theoretical foundation

Framework 4: Cross-Level Interaction

- Does some attribute of the higher-level entity impact the strength of the within-entity relationship?
  - Is the within-person relationship between leadership and depression stronger for individuals who have a heavy workload?

Framework 4: Cross-Level Interaction

- HRS.PM as moderator of Leadership – Depression Relationship

ID	TIME	DEPRESS	LEAD	HRS.PM
1	0	1.8	3	9.67
1	1	1.8	3	9.67
1	2	3.8	3	9.67
38	0	3.4	4	10.67
38	1	2.8	3	10.67
38	2	2.8	3	10.67
51	0	2.2	2	13.33
51	1	2	3	13.33
51	2	1.6	3	13.33

Framework 4: Cross-Level Interaction

- Cross-level interaction in an OLS fixed-effects framework
  - Has no main effect for HRS.PM only interaction
  - Also has 742 dummy codes for ID (not shown below)

```
> tmod<-lm(DEPRESS~LEAD+factor(TIME)+factor(ID)+LEAD:HRS.PM,
  CARMA2019)
> round(summary(tmod)$coef[c(1:4,747),],dig=3)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.123      0.444   7.039    0.000
LEAD          -0.246      0.160  -1.539    0.124
factor(TIME)1 -0.076      0.039  -1.931    0.054
factor(TIME)2 -0.210      0.040  -5.308    0.000
LEAD:HRS.PM    0.006      0.013   0.460    0.646
```

McNeish, D., & Kelley, K. 2019. Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, 24:20-35.

Framework 4: Cross-Level Interaction

- Cross-level interaction in a random effects model
  - Has a main effect for HRS.PM
  - Does not include 742 dummy codes for each individual

```
> tmod<-lme(DEPRESS~LEAD.DE*HRS.PM+factor(TIME),random=~1|ID,
  CARMA2019)
> round(summary(tmod)$tTable,dig=3)
                 Value Std.Error   DF t-value p-value
(Intercept)      1.734     0.225 1482   7.699   0.000
LEAD.DE         -0.246     0.160 1482  -1.539   0.124
HRS.PM           0.102     0.019  741   5.437   0.000
factor(TIME)1   -0.076     0.039 1482  -1.931   0.054
factor(TIME)2   -0.210     0.040 1482  -5.308   0.000
LEAD.DE:HRS.PM   0.006     0.013 1482   0.460   0.646
```

Framework 4: Cross-Level Interaction

- In a random-effect model we can formally test if the within slope between leadership and depression varies across individuals.
  - Person-means for work hours does not explain slope, but some other person-level predictor probably does

```
> tmod<-lme(DEPRESS~LEAD.DE+HRS.PM+factor(TIME),random=~1|ID,
  CARMA2019)
> tmod2<-update(tmod,random=~LEAD.DE|ID)
> anova(tmod,tmod2)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
tmod      1  7 5949.728 5989.678 -2967.864
tmod2     2  9 5939.886 5991.249 -2960.943 1 vs 2 13.84219   0.001
```

Framework 4: Cross-Level Interaction

- A clear hypotheses would state something like
  - The strength of the within-person relationship between leadership and depression varies by the average number of hours the person works such that…..

Framework 5: Growth Model

- The final variant moves TIME from being a source of unexplained heterogeneity to a variable of substantive interest

- Does the outcome increase or decrease over time?

Framework 5: Growth Model

- Growth models typically conducted in a random effects framework
  - Depression decreases over time (linear)
  - Could investigate quadratic trend as well in this example

```
> tmod<-lme(DEPRESS~TIME,random=~1|ID,CARMA2019)
> round(summary(tmod)$tTable,dig=3)
             Value Std.Error   DF t-value p-value
(Intercept)  2.915     0.036 1485  81.650   0.000
TIME        -0.068     0.020 1485  -3.475   0.001
```

Framework 5: Growth Model

- A key step is to investigate slope variability for higher-level entities
  - Model fits better if we allow each person to have a different slope

```
> tmod<-lme(DEPRESS~TIME,random=~1|ID,CARMA2019)
> tmod2<-update(tmod,random=~TIME|ID)
> anova(tmod,tmod2)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
tmod      1  4 6023.358 6046.192 -3007.679
tmod2     2  6 6013.622 6047.872 -3000.811 1 vs 2 13.73643   0.001
```

Framework 5: Growth Model

- Empirical Bayes' estimates of intercepts and slopes for 10 individuals

```
> coef(tmod2)[1:10,]
    (Intercept)        TIME
1      2.429283  0.15474563
38     3.071208 -0.11455317
51     2.306449 -0.14576257
55     2.903621  0.01499017
111    3.473553 -0.12086412
129    2.784424 -0.12625669
214    2.422009 -0.16377705
244    3.791807  0.07396698
311    3.015544 -0.16228565
327    2.872151 -0.16813741
```

Framework 5: Growth Model

- Person-means for leadership **do not** explain individual differences in the depression slopes

```
> tmod3<-lme(DEPRESS~TIME*LEAD.PM,random=~TIME|ID,CARMA2019)
> round(summary(tmod3)$tTable,dig=3)
                 Value Std.Error   DF t-value p-value
(Intercept)      4.014     0.150 1484  26.814   0.000
TIME            -0.087     0.091 1484  -0.950   0.342
LEAD.PM         -0.358     0.048  741  -7.540   0.000
TIME:LEAD.PM     0.006     0.029 1484   0.208   0.836
```

Framework 5: Growth Model

- Treating TIME as a vector is not a good way to recover the gold-standard estimate
  - Close to -0.173 but not as good as treating time as a factor

```
> tmod<-lme(DEPRESS~TIME+LEAD+LEAD.PM,random=~1|ID,CARMA2019)
> round(summary(tmod)$tTable,dig=3)
             Value Std.Error   DF t-value p-value
(Intercept)  4.033     0.126 1484  32.072       0
TIME        -0.105     0.020 1484  -5.333       0
LEAD        -0.175     0.021 1484  -8.173       0
LEAD.PM     -0.178     0.045  741  -3.959       0
```

Framework 5: Growth Model

- Fixed-effects model can estimate growth model interaction
  - No main effect for LEAD.PM
  - 742 dummy coded values not shown

```
> tmod<-lm(DEPRESS~TIME+factor(ID)+TIME:LEAD.PM,CARMA2019)
> round(summary(tmod)$coef[c(1:2,745),],dig=3)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.535      0.438   5.794    0.000
TIME            -0.087      0.086  -1.008    0.314
TIME:LEAD.PM     0.006      0.027   0.220    0.826
```

Framework 5: Growth Model

- For random-effect and fixed-effect models to provide identical estimates, the random effect model must have only a random intercept

```
> tmod<-lme(DEPRESS~TIME*LEAD.PM,random=~1|ID,CARMA2019)
> round(summary(tmod)$tTable,dig=3)
              Value Std.Error   DF t-value p-value
(Intercept)   4.014     0.151 1484  26.572   0.000
TIME         -0.087     0.086 1484  -1.008   0.314
LEAD.PM      -0.358     0.048  741  -7.472   0.000
TIME:LEAD.PM  0.006     0.027 1484   0.220   0.826
```

Framework 5: Growth Model

- Clear hypotheses in growth models would state something like
  - Depression will increase (or decrease) over time
  - Individuals will show significant variation in rates of increase (or decrease)
  - An individual's average rating of leadership will moderate the increase (or decrease) in depression such that….

Summary

- Purpose
- Panel Data
- Framework 1: Within-Effects
- Framework 2: Between-Effects
  - Inferences and Fallacies
- Framework 3: Incremental or Emergent Models
- Framework 4: Cross-level Interactions
- Framework 5: Growth Models
- Summary
- Conclusion

Conclusion

- The analysis of panel data can be approached in a variety of ways
  - No “right way”. Depends on theoretical question of interest and inference to be drawn
  - Do **not** recommend trying all approaches and seeing what works....

- Different disciplines tend to use different analytic approaches (fixed-effects versus random effects), but approaches can provide identical results

- Bliese et al., (in press) reviews published panel data in macro and micro literature and provides more detailed recommendations

UNIVERSITY OF SOUTH CAROLINA
Darla Moore School of Business

Back-Ups

Exogeneneity Defined

- We are somewhat forced into our broad research question because in an IDEAL statistical world
  - Participants would be randomly assigned to poor leaders or good leaders (CONDITION)
- Model would be exogenous
  - CONDITION unrelated to all other factors in model residuals

ID	CONDITION	DEPRESS
1	1	3.5
2	1	4.0
3	1	3.2
4	0	2.1
5	0	2.2
6	0	3.0

Exogeneity Defined

- In this design we could state that leaders *cause* depression
- Exogeneity is critical for establishing causality
  - Other possible causes are eliminated