

PANEL DATA ANALYSIS

Donald J. (DJ) Schepker

Associate Professor

Research Director, Center for Executive Succession

Moore Distinguished Fellow



UNIVERSITY OF
South Carolina

Darla Moore School of Business

**WHAT IS SOMETHING YOU ARE
HOPING TO TAKE AWAY FROM OUR
DISCUSSION TODAY?**



Darla Moore School of Business

PANEL DATA FEATURES AND ADVANTAGES

- Multiple observations of a given entity over time (e.g. firms, people, industries)
 - Sometimes referred to as longitudinal data and can be applied in cases like experience sampling methodologies (ESM)
- Advantages of Panel Data
 - Explores variance both across entities (cross-sectional variation) and over time (temporal variation)
 - Captures unobserved heterogeneity (e.g. firm, person)
 - By combining across time and space, there is greater variation, less collinearity, and more degrees of freedom
 - Can detect effects not observed in cross-section or time-series data

DATA STRUCTURE

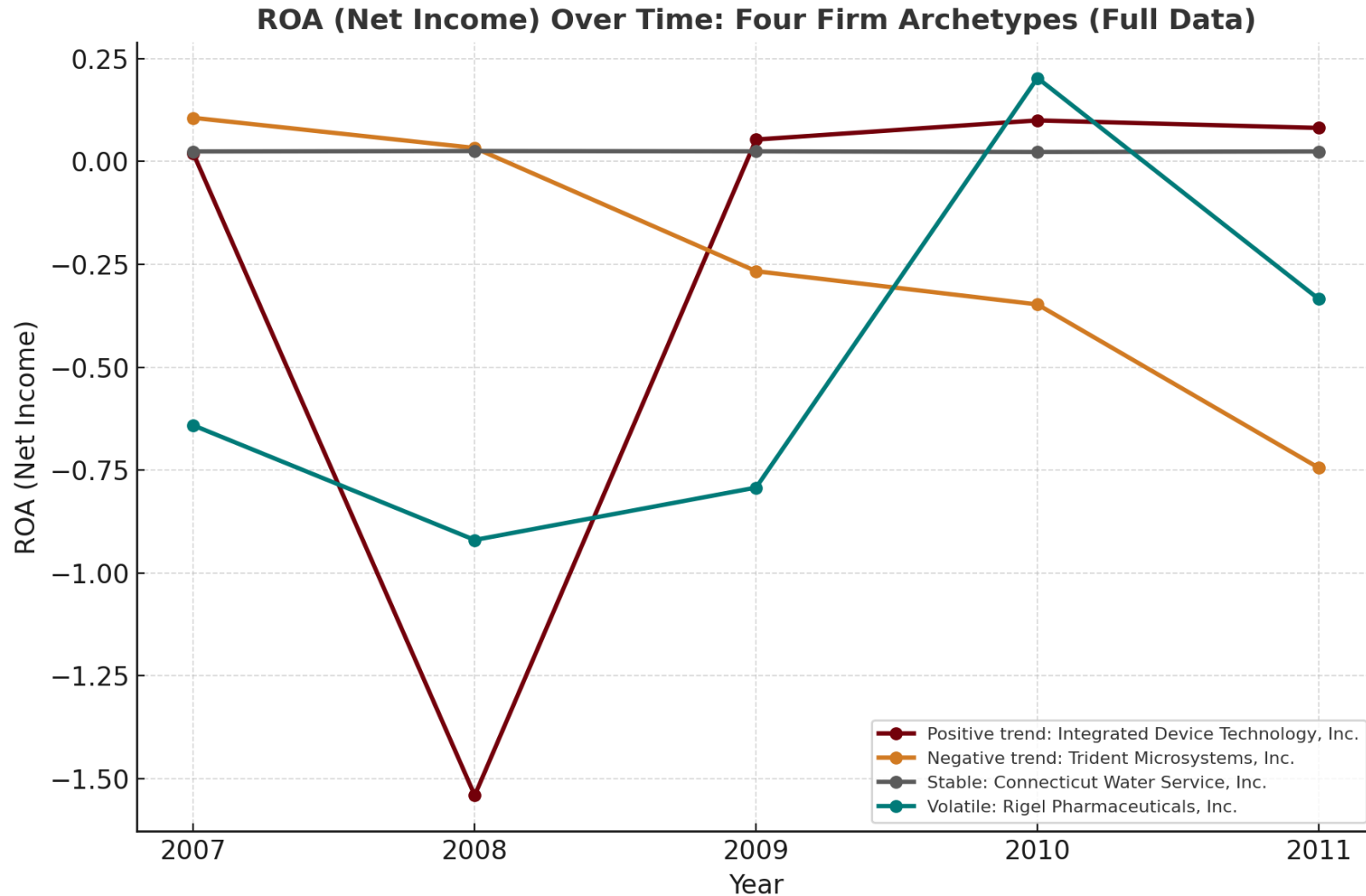
- Data typically entered by both entity and year
 - n = number of entities (e.g. people, firms, industries)
 - t = number of years
- Each observation or record typically relates to entity-year
 - Balanced panel: same number of observations for each entity
 - Unbalanced panel: entities are represented by different numbers of observations
- Key variables
 - ID: variable which represents each unique entity
 - Time: variable representing each observation; important to have distance between time periods equidistant
 - Dependent variable: outcome of interest
 - Predictors: variables of interest

VISUALIZING THE DATA

ID	Company	year	roani	BoardIndProp	DirAgeAvg	WomenDirs
568	DSP Group, Inc.	2007	-0.0093	0.8333	53.1667	0.0000
568	DSP Group, Inc.	2008	-0.8521	0.8333	54.1667	0.0000
568	DSP Group, Inc.	2009	-0.0384	0.8889	52.1111	0.0000
568	DSP Group, Inc.	2010	-0.0334	0.8750	50.0000	0.0000
568	DSP Group, Inc.	2011	-0.0828	0.7500	51.1000	0.0000
2245	Rigel Pharmaceuticals, Inc.	2007	-0.6414	0.0000	57.0000	0.0000
2245	Rigel Pharmaceuticals, Inc.	2008	-0.9200	0.7778	57.6667	0.0000
2245	Rigel Pharmaceuticals, Inc.	2009	-0.7926	0.7778	58.3333	0.0000
2245	Rigel Pharmaceuticals, Inc.	2010	0.2030	0.7778	57.9000	0.0000
2245	Rigel Pharmaceuticals, Inc.	2011	-0.3344	0.7778	58.9000	0.0000
2360	Select Comfort Corporation	2007	0.1450	0.9000	54.9000	0.2000
2360	Select Comfort Corporation	2008	-0.5182	0.9000	55.9000	0.2000
2360	Select Comfort Corporation	2009	0.3007	0.9091	55.8182	0.2727
2360	Select Comfort Corporation	2010	0.1857	0.9000	57.4545	0.2000
2360	Select Comfort Corporation	2011	0.2303	0.8750	58.7778	0.1250
2689	Trident Microsystems, Inc.	2007	0.1061	0.8000	68.6000	0.0000
2689	Trident Microsystems, Inc.	2008	0.0328	0.6667	73.5000	0.0000
2689	Trident Microsystems, Inc.	2009	-0.2667	0.8571	59.5714	0.1429
2689	Trident Microsystems, Inc.	2010	-0.3475	0.7143	59.1250	0.1429
2689	Trident Microsystems, Inc.	2011	-0.7447	0.4444	57.2000	0.1111
2871	Warren Resources, Inc.	2007	0.0259	0.5556	61.5556	0.0000
2871	Warren Resources, Inc.	2008	-0.8427	0.6000	63.8000	0.0000
2871	Warren Resources, Inc.	2009	-0.0527	0.6000	63.8000	0.0000
2871	Warren Resources, Inc.	2010	0.0748	0.6667	63.3000	0.0000
2871	Warren Resources, Inc.	2011	0.0669	0.6667	64.2000	0.0000

- Substantial within-entity variation (roani)
- Less within variation in board independence and director average age, but greater between
- Both within and between variation for women directors

VISUALIZING THE DATA



KEY CONSIDERATIONS: SOURCES OF VARIATION

- In knowing your data, it is important to understand where variation lies
- Intra-class correlation coefficient (ICC)
 - Reveals how much of the outcome is driven by entity-level vs. time-level factors
 - Helps decide which model structure fits best
 - Essential for partitioning variance before estimating complex models
 - Bounded between 0 and 1
 - If $ICC(1)$ is 0.72, then 72% of variation in the outcome occurs between entities
 - Conditional ICC
 - Illustrates how much variance remains at each level after accounting for specific predictors, such as time
 - Entity and time conditional ICCs quantify remaining clustering after absorbing entity and time effects

Shows where the action is and helps justify model choice



**Darla Moore
School of Business**

KEY CONSIDERATIONS – N VS. T

- Do we want large n or large t?
 - Ideally, both!
 - If interested in between-entity variation, large n
 - For within-entity variation, large t
 - Having low t can potentially bias inferences from pooled data; however, this can be limited with clustered standard errors
- Quick exercise:
 - What is a research question you currently have or might have that could be explored with panel data?
 - What would the data structure look like? What is the ID variable? Time variable? Dependent variable?

ANALYTICAL METHODS

- Pooled OLS
- Random Effects
- Fixed Effects

ANALYTICAL METHODS – POOLED OLS

- Treats every observation as independent, even when the same entity appears over multiple years
- Assumptions
 - All observations are independent and identically distributed
 - No correlation across time within a firm
- Ignores
 - Unobserved entity effects
 - Serial correlations
- Implications
 - Biased coefficients if unobserved entity traits correlate with x
 - Standard errors likely too small (Type II error)
- Most acceptable when:
 - Minimal within-entity correlation
 - Short panels with little repeated measurement

Pooled OLS - Great for intuition; risky for inference.



**Darla Moore
School of Business**

ANALYTICAL METHODS – RANDOM EFFECTS

- Assumes firm-specific effects are random and uncorrelated with the regressors, combining within- and between-firm information.
- Concept
 - Treats unobserved entity characteristics as a random variable uncorrelated with predictor variables
 - Uses both within- and between-entity variation for estimation
- Interpretation
 - Coefficients reflect an average effect across entities and over time
- Advantages
 - More efficient if assumptions hold
 - Allows estimation of time-invariant variables (e.g. industry)
- Limitations
 - Biased if random firm effects are correlated with the predictors of interest
 - Requires Hausman test to compare with Fixed Effects specifications

Combines within and between information — efficient but assumption-sensitive.



**Darla Moore
School of Business**

ANALYTICAL METHODS – FIXED EFFECTS

- Controls for all firm-specific, time-invariant characteristics by using only within-firm variation.
- Concept:
 - Removes each firm's average outcome and predictors (dumming or demeaning)
 - Focuses on how changes within an entity over time relate to changes in Y
- Advantages:
 - Eliminates bias from unobserved, stable firm characteristics
 - Preferred when individual effects correlate with regressors
- Limitations:
 - Cannot estimate effects of time-invariant variables
 - Less efficient if unobserved entity specific effects are uncorrelated with predictors
 - Within-firm variation must be sufficient

Within-Firm Analysis – isolates entity changes over time.



**Darla Moore
School of Business**

HAUSMAN TEST

- Test of consistency in coefficient estimates across random vs. fixed effects specifications
- Concept
 - Examines the coefficients in the random effects model vs. those in the fixed effects models
 - Null hypothesis: Random effects is consistent and efficient
 - Alternative: random effects is inconsistent; use fixed effects
- Limitations
 - Can identify presence of unobserved heterogeneity, but not how to deal with it
 - Omnibus test across all predictors estimated; can be sensitive to model size and multicollinearity
- Most reliable for parsimonious models; consider testing by subset

WHAT SHOULD DRIVE THE CHOICE OF WHICH MODEL TO USE?



Darla Moore School of Business

SELECTING THE APPROPRIATE MODEL

- Conceptually – is heterogeneity related to the predictors of interest?
- Methodologically
 - Examine the efficiency vs. consistency tradeoff by performing a Hausman test
 - Understand where variation lies in the data
- Theoretically
 - Is the research question focused on between-entity variation, within-entity variation, or both?

PANEL DATA – AN ANALYTICAL ILLUSTRATION

- Random draw of 196 publicly traded firms from 2008-2012
- Balanced panel (5 years of observations)
- DV = Firm ROA expressed as a percentage
- IVs
 - Board Independence – proportion of directors not affiliated with the firm
 - Average age of directors on the board
 - Number of women directors on the board
- All data and code for analyses presented can be found here:
https://osf.io/dgvh7/?view_only=fe7b3c79fc9043a6a730c677941b0de4

UNDERSTANDING SOURCES OF VARIATION

- ICC – estimates total variance that can be attributed to the higher-level entities using a null model
 - Adjusted r-square from a fixed-effects model only including firms
- ICC for ROA is 0.35 – what does this mean?
 - 35% of variance is attributable to firm-level differences (65% of variance can be found within a firm)
- Conditional ICC based on entity and time – 0.3594
 - About 36% of variance exists between firms even when controlling for time effects, 64% remains between firms

AN ILLUSTRATION – POOLED OLS

DV=ROA	Coeff	Std Error	P-value
year			
2008	-4.35	(1.37)	0.00
2009	-2.72	(0.79)	0.00
2010	-0.22	(0.87)	0.80
2011	-0.79	(0.96)	0.41
Board Independence	-1.81	(3.92)	0.65
Directors Average Age	0.02	(0.18)	0.92
Women Directors	18.75	(7.27)	0.01
Intercept	3.99	(11.37)	0.73
<i>Observations</i>		980	

How would we interpret these results?

AN ILLUSTRATION – RANDOM EFFECTS

DV=ROA	Coeff	Std Error	P-value
year			
2008	-4.52	(1.07)	0.00
2009	-2.73	(1.08)	0.01
2010	-0.19	(1.07)	0.86
2011	-0.98	(1.08)	0.36
Board Independence	-0.79	(3.64)	0.83
Directors Average Age	0.30	(0.15)	0.05
Women Directors	15.53	(6.22)	0.01
Intercept	-15.40	(9.39)	0.10
<i>Observations</i>		980	
<i>Groups</i>		196	

How would we interpret these results?

AN ILLUSTRATION – FIXED EFFECTS

DV=ROA	Coeff	Std Error	P-value
year			
2008	-4.91	(1.07)	0.00
2009	-2.69	(1.08)	0.01
2010	-0.08	(1.08)	0.94
2011	-1.41	(1.11)	0.20
Board Independence	0.21	(4.74)	0.97
Directors Average Age	1.12	(0.25)	0.00
Women Directors	9.74	(10.78)	0.37
Intercept	-64.59	(15.31)	0.00
<i>Observations</i>		980	
<i>Groups</i>		196	

Incorporating fixed effects provides a clean firm-year (within-firm) estimate

Interpretation

Within firms, ROA is higher when the average age of directors is higher.

AN ILLUSTRATION – MIXED EFFECTS WITH DUMMIES

DV=ROA	Coeff	Std Error	P-value
year			
2008	-4.91	(0.95)	0.00
2009	-2.69	(0.96)	0.01
2010	-0.08	(0.96)	0.93
2011	-1.41	(0.99)	0.15
Board Independence	0.21	(4.22)	0.96
Directors Average Age	1.12	(0.22)	0.00
Women Directors	9.74	(9.60)	0.31
Intercept	-69.00	(15.01)	0.00
<i>Observations</i>	980		
<i>Groups</i>	196		

GLS and Multilevel modeling (mixed effects) provide same fixed effects parameter estimates, but are those estimates different from the previously illustrated random effects estimates?

COMPARISON OF RESULTS

	Pooled OLS		Random Effects		Fixed Effects	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Board independence	-0.79	0.65	-0.79	0.83	0.21	0.97
Directors' Average Age	0.30	0.92	0.30	0.05	1.12	0.00
Women Directors	15.53	0.01	15.53	0.01	9.74	0.37

- In econometrics, next step is run the Hausman test – significant results indicate random effects estimates are “polluted”
- In micro research, results suggest emergent effects
 - Cautious against making inferential fallacy (atomistic or ecological)

HAUSMAN TEST

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) fixed	(B) random		
year				
2008	-4.906628	-4.518087	-.388541	.0074927
2009	-2.693257	-2.731603	.0383463	.1188086
2010	-.0795868	-.1941232	.1145365	.1097883
2011	-1.409539	-.9785024	-.4310367	.2313508
BoardIndep~p	.2077942	-.790766	.9985602	3.039872
DirAgeAvg	1.124924	.3017773	.823147	.198144
FemaleDirs~p	9.742956	15.52639	-5.783437	8.807401

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
 = 21.48
 Prob>chi2 = 0.0031
 (V_b-V_B is not positive definite)

- Coefficients are not consistent across the two models

Suggests that unobserved firm heterogeneity is driving differences in results

But is indicative of emergent effects across levels (but does not specify where)

ICC suggests that there is considerable variance to explain both within and between firms, while the Hausman test suggests that the nature of the specific relationships differs across levels

ANALYTICAL METHODS – HYBRID MODEL

- Allison (2009) outlines a hybrid model to explore within and between variation
- Concept:
 - Includes a variable for each entity's average for each predictor (group mean)
 - Incorporates variables for the raw value of predictors (tests of emergence) or demeaned values (group-mean centering; within-entity variation)
- Advantages:
 - Eliminates bias from unobserved, stable characteristics on the predictors of interest
 - Able to parse variance into within and between effects
 - Estimates effects for time-invariant variables
- Limitations:
 - Can be biased if random effects are correlated with predictors
 - Likely requires adjusting standard errors for homoskedasticity and non-independence

TEST OF EMERGENCE IN RANDOM EFFECTS MODELS

DV=ROA	Coeff	Std Error	P-value
year			
2008	-4.91	(1.07)	0.00
2009	-2.69	(1.08)	0.01
2010	-0.08	(1.08)	0.94
2011	-1.41	(1.11)	0.20
Board Independence	0.21	(4.74)	0.97
Directors Average Age	1.12	(0.25)	0.00
Women Directors	9.74	(10.78)	0.37
Board Independence Mean	-2.99	(7.29)	0.68
Directors Average Age Mean	-1.33	(0.32)	0.00
Women Directors Mean	11.18	(13.19)	0.40
Intercept	15.80	(11.85)	0.00
Observations		980	
Groups		196	

- How would you interpret these coefficients?

TEST OF EMERGENCE IN RANDOM EFFECTS MODELS

DV=ROA	Coeff	Std Error	P-value
year			
2008	-4.91	(1.07)	0.00
2009	-2.69	(1.08)	0.01
2010	-0.08	(1.08)	0.94
2011	-1.41	(1.11)	0.20
Board Independence	0.21	(4.74)	0.97
Directors Average Age	1.12	(0.25)	0.00
Women Directors	9.74	(10.78)	0.37
Board Independence Mean	-2.99	(7.29)	0.68
Directors Average Age Mean	-1.33	(0.32)	0.00
Women Directors Mean	11.18	(13.19)	0.40
Intercept	15.80	(11.85)	0.00
Observations		980	
Groups		196	

- Coefficient for Directors average age suggests that the relationship differs between levels (within vs. between levels)
- *The between firm effect is 1.33 units less than the within firm effect*

THE HYBRID MODEL – DEMEANING

DV=ROA	Coeff	Std Error	P-value
year			
2008	-4.91	(1.07)	0.00
2009	-2.69	(1.08)	0.01
2010	-0.08	(1.08)	0.94
2011	-1.41	(1.11)	0.20
Board Independence Demean	0.21	(4.74)	0.97
Directors Average Age Demean	1.12	(0.25)	0.00
Women Directors Demean	9.74	(10.77)	0.37
Board Independence Mean	-2.78	(5.58)	0.62
Directors Average Age Mean	-0.20	(0.19)	0.29
Women Directors Mean	20.92	(7.63)	0.01
Intercept	15.80	(11.91)	0.00
Observations		980	
Groups		196	

- How would you interpret these coefficients?

THE HYBRID MODEL – DEMEANING

DV=ROA	Coeff	Std Error	P-value
year			
2008	-4.91	(1.07)	0.00
2009	-2.69	(1.08)	0.01
2010	-0.08	(1.08)	0.94
2011	-1.41	(1.11)	0.20
Board Independence Demean	0.21	(4.74)	0.97
Directors Average Age Demean	1.12	(0.25)	0.00
Women Directors Demean	9.74	(10.77)	0.37
Board Independence Mean	-2.78	(5.58)	0.62
Directors Average Age Mean	-0.20	(0.19)	0.29
Women Directors Mean	20.92	(7.63)	0.01
Intercept	15.80	(11.91)	0.00
Observations		980	
Groups		196	

- Note that the difference between the demeaned coefficient and the mean coefficient was the mean coefficient in the prior model
 - Age = 1.12-1.33 = 0.20 (rounding)
- Directors' average age mean coefficient does not illustrate significant between firm difference from 0
- However, firms with higher levels of average women directors have higher average ROA

CAUTIONS AND IMPORTANT ISSUES TO IDENTIFY

- Unbalanced panel data (in general mixed-effects are designed to handle unbalanced)
 - Fixed effects estimates are underlied by ANOVA models, making unbalanced panel data estimates slightly more unreliable if the missing data is owed to sample selection (not missing at random)
- Invariance in predictor variables
 - When control variables or independent variables have small variance over time (e.g. firm size), inclusion in a fixed effect model may be problematic
- Hausman test is often an omnibus test of all coefficients, rather than only the predictors of interest
- The hybrid model only controls for unobserved heterogeneity for predictors which have group-level means entered

PANEL DATA EXERCISE

- Considering the research questions outlined earlier...
 - Where would you expect variation to lie in the data?
 - Which model would you choose to run?
 - Define the model's specification
 - How would you interpret the results?

CONSIDERATIONS OF TIME

- Many strategy models incorporate time fixed effects to limit the influence of unobserved heterogeneity over time
- This relegates time to a variable to be controlled for, rather than one which has explanatory power
- Micro researchers use a variety of growth models that might help provide explanatory power in our research

GROWTH MODELS – BASIC MODEL

DV=ROA	Coeff	Std Error	P-value
Debt to Assets	-6.83	(2.72)	0.01
Firm size	1.61	(0.33)	0.00
CEO Tenure	0.12	(0.07)	0.09
Time	0.24	(0.24)	0.31
Intercept	-0.30	(1.22)	0.80
<i>Observations</i>		980	
<i>Groups</i>		196	

- Includes time as a linear variable, rather than controlling for each period as a dummy variable
 - Both could be done if all firms are not sampled in the same periods
- Result - ROA is not significantly changing over time (in a linear fashion)

PREDICTORS INFLUENCING CHANGE OVER TIME

DV=ROA

	Coeff	Std Error	P-value
Debt to Assets	-7.15	(2.72)	0.01
Firm size	1.69	(0.33)	0.00
CEO Tenure	0.13	(0.07)	0.07
Time	10.75	(4.22)	0.01
Directors Avg Age Mean	-0.09	(0.23)	0.69
Dir Avg Age Mean x time	-0.18	(0.07)	0.01
Intercept	5.03	(13.86)	0.72

Observations

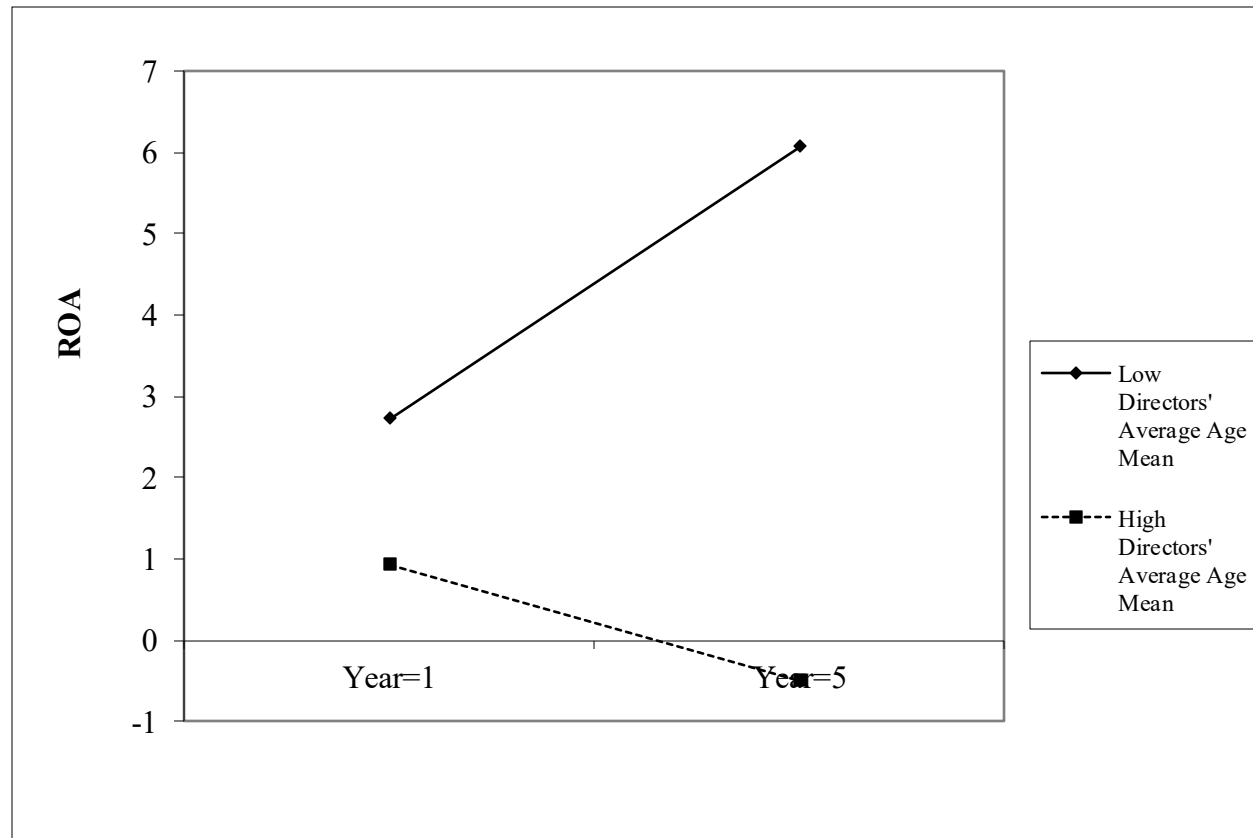
980

Groups

196

- Incorporate a predictor interaction with time
- Result – Average age of directors has a negative relationship to the growth of ROA

ILLUSTRATION OF THE GROWTH TREND



- Firms that had higher mean levels of the directors' average age over the 5 years saw a decline in ROA, while those with lower directors' average age over the sample window saw ROA growth

PANEL DATA EXERCISE – CONSIDERING TIME

- If we have time...
- Again considering the research questions outlined earlier...
 - How might the outcome of interest change over time?
 - Are there entity-level predictors that might influence the trajectory over time?

CONCLUSION

- Use theory to drive model choice!!!!
- Fixed effect models are often the default with panel data, however, they have severe limitations
 - Test within effects only
 - May run into issues with invariant predictor variables
 - Underlied by ANOVA models making more unreliable if unbalanced panel data is owed to data not missing at random
- Clustered standard errors do not account for unobserved firm heterogeneity, only if observations within entity are not i.i.d
- Test for emergent effects within and between entities, and use ICCs to examine the variation to be explained
- Harness the explanatory power of time, particularly in multi-level models
 - Regardless, the role of “time” needs to be methodologically considered and likely modeled in some fashion

REFERENCES / RESOURCES

- 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(7), 1536-1556.
- Bliese, P. D., Schepker, D. J., Essman, S. M., & Ployhart, R. E. (2020). Bridging methodological divides between macro-and microresearch: Endogeneity and methods for panel data. *Journal of Management*, 46(1), 70-99.
- Schepker, D. J., & Bliese, P. D. (2023). Using mixed-effect growth models to examine time as a predictor of interest and between-firm differences over time. In *Methods to Improve Our Field* (Vol. 14, pp. 79-99). Emerald Publishing Limited.
- Shaver, J. M. (2019). Interpreting interactions in linear fixed-effect regression models: When fixed-effect estimates are no longer within-effects. *Strategy Science*, 4(1), 25-40.

THANKS!

Donald J. (DJ) Schepker
Associate Professor, Management
Distinguished Moore Fellow
Research Director, Center for Executive Succession
DJ.Schepker@moore.sc.edu
<http://www.linkedin.com/in/DJSchepker>



**Darla Moore
School of Business**