Class 13 - Design II - Longitudinal

Agenda

- Longitudinal designs: what, why, when, how (30 minutes)
- Application paper discussion (30 minutes)
- Replication presentation (20 minutes)
- Skills corner - Class walkthrough in R (25 minutes)
- General discussion (10 minutes)
- *Break* then concluding remarks

Longtudinal designs

What is a longitudinal design?

Taking all the above into consideration, we define longitudinal research as research emphasizing the study of change and containing at minimum three repeated observations (although more than three is better) on at least one of the substantive constructs of interest. (Ployhart and Vandenberg 2010, 97)

Why are longitudinal designs useful?

[S]trategy researchers have not fully capitalized on the fact that longitudinal data are multilevel in nature. Longitudinal data include two types of variance: within-unit (e.g., within-firm, within-person, etc.) and between-unit (e.g., between-firm, between-person, etc.) variance1. These two types of variance correspond to two different relationships: within-firm and between-firm relationships. (Certo, Withers, and Semadeni 2017, 1536)

How can you design a longitudinal study?

A multitude of potential designs

- Growth curve modeling (Ployhart and Vandenberg 2010)
- Moderation by causal cycle (Mitchell and James 2001)
- Panel methods (Certo, Withers, and Semadeni 2017)
- Dynamic panel models
- Dynamic SEM models (e.g., cross-lagged panel models)
- Time series modeling (ARIMA, GARCH)
- Markov chain monte carlo (MCMC)
- Stochastic differential equations (SDEs)

A basic workhorse: Panel data estimation

Panel data methods exploit within and between entity variance, and blend some of the benefits of cross-sectional and time series designs

Fixed effects estimators are also consistent in the presence of unmeasured or unobservable confounding variables that are **time invariant** for the period of analysis (e.g., general mental ability, firm quality, etc.)

A basic workhorse: Panel data estimation

$$Y_{it} = \beta X_{it} + \nu + \epsilon$$

There are basically two key questions in a typical panel data setup:

- Can the data across time be pooled (essentially is $Var(\nu) = 0$)?
- Is the idiosyncratic error for each entity uncorrelated with the regressors (is $E(\nu|X) = 0$)?

If the first is true, we can used "pooled OLS". If the second is true, we can use "random effects". Otherwise, we need to use "fixed effects".

Three ways to get to a fixed effects estimator

- Include a dummy variable for each entity
- The "within" transformation - subtract the mean from every observation
- First differencing - take a one period difference between each observation

They will typically provide the same answer but there can be differences (particularly for the first difference operator) if there are multiple panel waves and the changes over time fluctuate about the overall average

An alternative workhorse model: Lagged dependent variables

$$Y_{it} = \beta X_{it} + \theta Y_{it-1} + \epsilon$$

Here, the model can be interpreted two ways:

- X predicts the incremental change in Y
- Y is predicted by its past history, which means:
    - all of the X variables have a multiplier effect proportional to the coefficient on Y, and
    - the error term likewise propogates through time

(Note it is also possible to have a model where errors are autocorrelated but there are no dynamics) (Keele and Kelly 2006)

Why not just use fixed effects and lagged DVs?

In short, the assumptions conflict.

The conditions for consistent estimation of this model are much more demanding than those required with fixed effects or lagged dependent variables alone. [...] The problem here is that the differenced residual is necessarily correlated with the lagged dependent variable because both are a function of the fixed effect in the prior period (Angrist and Pischke 2008, 183).

There are lagged panel models that incorporate instrumental variables, but they chew up several years and have the typical issues of small-sample bias.

Not all hope is lost

> *[W]hat ís an applied guy to do? One answer, as always, is to check the robustness of your findings using alternative identifying assumptions. That means that you would like to find broadly similar results using both models. Fixed effects and lagged dependent variables estimates also have a useful bracketing property. (Angrist and Pischke 2008, 184)*

Also, with a long enough time panel (as $T \to \infty$), the magnitude of the Nickell bias becomes smaller (Baum 2013) (but its not a perfect solution since convergence is $O(1/T)$).

Why I am focusing on these two models?

Between them, they are able to address many of the endogeneity problems discussed in Week 11

They also capitalize on within- and between- entity variation and can be generalized using tools like RCM and the like

The drawback is their continued grounding in a regression-type framework which may not fully capture some of the nuances of longitudinal study or at a minimum impose some substantial assumptions

Quick note on terminology

Panel data econometrics uses different terminology than statistics and other social sciences. Here is the (loose) translation guide.

Model	Econometrics	Statistics
$Y_{it} = \beta X_{it} + \alpha_i + \epsilon_{it}$	Fixed effects	LSDV (Least squares dummy variable)
$Y_{it} = \beta X_{it} + \nu + \epsilon_{it}$	Random effects	Fixed effects
$Y_{it} = \beta_i X_{it} + \nu + \epsilon_{it}$	Random coefficient modeling (RCM)	Random effects

Confusing I know... this is what happens when different fields derive similar methods independently.

Applications

Application readings

Let's level-set people's familiarity with these pieces.

- 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.
- Firm Repertoires and Performance: The Influence of Complementarity and Competition (working paper)

Certo Withers and Semadeni (2017)

- What was this paper about?
- What were the findings?
- What was the method?
- What makes sense? What was confusing?

Strategic Management Journal
Strat. Mgmt. J., **38**: 1536–1556 (2017)
Published online EarlyView 25 October 2016 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2586
Received 21 November 2015; Final revision received 16 August 2016

A TALE OF TWO EFFECTS: USING LONGITUDINAL DATA TO COMPARE WITHIN- AND BETWEEN-FIRM EFFECTS

S. TREVIS CERTO,[1] MICHAEL C. WITHERS,[2] and MATTHEW SEMADENI[1,*]
[1] *W.P. Carey School of Business, Arizona State University, Tempe, Arizona, U.S.A.*
[2] *Mays Business School, Texas A&M University, College Station, Texas, U.S.A.*

Research summary: *We investigate the theoretical and empirical implications of longitudinal data in strategy research. Theoretically, longitudinal data allow strategy researchers to distinguish between relationships among constructs* within *versus* between *firms. Empirically, longitudinal data contain information about two types of relationships: within- and between-firm. We describe how the hybrid approach, a technique used in other disciplines, disentangles within- and between-firm relationships. We reexamine a study of research and development expenditures to illustrate the advantages of the hybrid approach. Based on our theory and reexamination, we offer a series of recommendations for researchers using longitudinal data to test theoretical perspectives.*

Managerial summary: *Strategy research examines two sources of variation over time: what is occurring within the firm (e.g., Do firms perform better over time when investing more in R&D?) and what is occurring between firms (e.g., Do firms investing more in R&D outperform firms investing less in R&D?). These two sources may be similar or different in both direction and magnitude, and when significant differences exist in either direction or magnitude, researchers must carefully consider the implication of these differences to their theoretical rationale and statistical testing. Our article highlights the benefits of theorizing and testing these two sources of variance, providing scholars the ability to broaden both the theoretical and empirical contribution of their research. This distinction is important to how research informs managerial decision making.*

INTRODUCTION

Since the early 2000s, strategy scholars have increasingly relied on longitudinal data to test theorized relationships. In 2004, approximately 15 percent of articles published in *SMJ* involved longitudinal data. By 2014, *more than half* of articles published in *SMJ* involved longitudinal data. During that period, researchers used longitudinal data to test a variety of topics, including sustainability strategies (e.g., Bansal, 2005), firm reputation (e.g., Basdeo *et al.*, 2006), merger waves (e.g., Haleblian *et al.*, 2012), strategic alliances (e.g., Koka and Prescott, 2008), downsizing (e.g., Love and Nohria, 2005), and CEO compensation (e.g., Bodolica and Spraggon, 2009).

Despite this remarkable increase, strategy researchers have not fully capitalized on the fact that longitudinal data are multilevel in nature. Longitudinal data include two types of variance: within-unit (e.g., within-firm, within-person,

Keywords: research methods; theory testing; longitudinal data; theory development; hybrid approach
*Correspondence to: Matthew Semadeni, PO Box 874006, Tempe AZ 87287-4006. E-mail: semadeni@asu.edu

Fox Simsek and Souder (working paper)

- What was this paper about?
- What were the findings?
- What was the method?
- What makes sense? What was confusing?

15401

Firm Repertoires and Performance: The Influence of Complementarity and Competition

ABSTRACT

Two perspectives exist on how firms should choose the set of actions, collectively known as a competitive repertoire, intended to generate superior returns. One is a position-based approach that emphasizes consistency and fit between capabilities and external demands. The other perspective argues such positions can be easily imitated, necessitating repertoires with higher complexity to keep competitors off-balance while also enabling higher performance through increased adaptability. This paper helps untangle whether and when repertoire complexity and consistency are beneficial for firm performance by examining their influence across varying levels of competitive intensity – a key determinant which heretofore has been under examined. Based on an examination of financial performance, we find that the optimum level of consistency decreases as competitor activity increases, while the performance enhancing effects of complexity manifest to a greater extent at high levels of competitor activity.

Keywords:

Competitive repertoires; temporary advantage; firm performance; sustainable competitive

Break

Replication Presentation

Replication: Firm Repertoires and Performance: The Influence of Complementarity and Competition (working paper)

Skills corner - Class walkthrough in R

Preparation for next class

Next class

- Concluding remarks
- Final paper working session

References

Angrist, J. D., and J. S Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Certo, S. Trevis, Michael C. Withers, and Matthew Semadeni. 2017. "A Tale of Two Effects: Using Longitudinal Data to Compare Within- and Between-Firm Effects." *Strategic Management Journal* 38 (7): 1536–56.

Mitchell, Terence R., and Lawrence R. James. 2001. "Building Better Theory: Time and the Specification of When Things Happen." *The Academy of Management Review* 26 (4): 530–47.

Ployhart, Robert E., and Robert J. Vandenberg. 2010. "Longitudinal Research: The Theory, Design, and Analysis of