

Class Schedule and Readings for Research Methods in Organization Science II

MGT 791 (Spring 2018)

ver: Jan 3, 2018

Setting the stage	2
Session 1 Theory and data analysis	2
Session 2 A review of Ordinary Least Squares	6
Session 3 More on OLS and current controversies	8
Bayesian analysis	13
Session 4 A brief look at Bayesian analysis	13
Repeat appearances: Multilevel and panel data	16
Session 5 Panel data 1	17
Session 6 Panel data 2 and Multilevel Models 1	19
Session 7 Multilevel 2	22
Limited dependent variables	24
Session 8 Logit and probit - basics and interpretation	25
Spring Break	28
Session 9 Logit and probit - complications and extensions	29
Session 10 Hazard rate models 1	32
Session 11 Hazard rate models 2	34
Session 12 Count data models	36
Failure of exogeneity	37
Session 13 Truncation and Censoring	38
Session 14 Sample selection	41
Session 15 Endogeneity	43
Conclusion and finals	46
Pseudo-session 16 Conclusion and finals	46

Setting the stage

Session 1 Theory and data analysis

Jan 9, 2017

Before we dive into any statistics, there are some important preliminaries to complete. First, we'll introduce ourselves and I'll introduce you the purpose and logistics of the class. Then we'll discuss the intersection of theory and method. After all, if we don't know the purpose of our research methodology, all the statistical sophistication in the world won't help us much. Then we'll talk about managing the data analysis process—an overall philosophy and practical ideas that can save you a lot of pain over the lifetime of your research projects. In addition to the readings (and one video) listed below, please be sure to complete the pre-class assessment as described under “Assignment”. You'll find everything in the “For Session 1” folder on Dropbox.

Reading

- Theory, method and their intersection

There are two conceptual clusters here. The first cluster you should explore addresses approaches to scientific research generally. Start by watching the Stearns video. Then read Chamberlin and Nelson. The second cluster of Aguinis & Edwards, Bettis, and Van Maanen *et al* is specific to management research. I would start with Van Maanen *et al* and then read the other two in whichever order you want.

- Aguinis H, Edwards JR. 2014. Methodological wishes for the next decade and how to make wishes come true. *Journal of Management Studies* **51**(1): 143-174

Aguinis and Edwards are both prominent scholars in the field of management research methodology, so it's worth knowing where they think the field should be going in the future. You can skim to some degree. Primarily focus on their seven wishes and, most importantly, why they think fulfilling the wishes is important to the field.

- Bettis RA. 2012. The search for asterisks: Compromised statistical tests and flawed theories. *Strategic Management Journal* **33**(1): 108-113

Bettis is a Senior Editor at *SMJ* and has always struck me as especially insightful. What methodological concern is he focusing on, why does it matter and how does he suggest improving our practice?

- Chamberlin TC. 1965. The method of multiple working hypotheses. *Science* **148**(3671): 754-759

Originally published in 1890. Please be ready to discuss whether you think the method of multiple working hypotheses is (a) valid and (b) applicable in your research.

- Nelson RR. 2015. Numbers and math are nice, but.... *Biological Theory* **10**(3): 246-252

Why not start a management seminar by reading a paper by an economist that appeared in a biology journal...Actually, Nelson has been as influential in strategy as in economics *per se*. As an economist, however, he's given a lot of thought to the role of formal models, which I've why I've chosen this piece. Please focus on (a) his comparison of the role of quantitative methods in physics versus other fields, (b) his view on the appropriate use of models, and (c) his thoughts about measurement. His thoughts on economic growth from an evolutionary viewpoint are well worth reading for their own sake, but you don't need to worry about them for our purposes.

- Stearns SC. 2009. Lecture 23 - the logic of science, *EEB 122: Principles Of Evolution, Ecology And Behavior*. Yale University. <http://oyc.yale.edu/ecology-and-evolutionary-biology/eeb-122/lecture-23>

From Yale University's class *Principles Of Evolution, Ecology And Behavior*. While the examples are from outside our field, it is a superb overview of how scientists *should* think. The URL I've provided gives you access to multiple ways of viewing this material. There is a transcript, but I think watching the video (especially the higher resolution version, so you can see the slides) is worthwhile.

- Van Maanen J, Sorensen JB, Mitchell TR. 2007. The interplay between theory and method. *Academy of Management Review* **32**(4): 1145-1154

Based on your experience and your reading of this paper, is it possible to—as they put it—“serve [the] two masters” of theory and method? Why or why not? If not, when and why should one dominate over the other?

- Managing the data analysis process

- Hoetker G. 2015. “Organizing Data, Do Files And Results” *Standard Operating Procedures*. Pages 8-17.

This is a document I share with my Research Assistants. It's under pretty constant revision, so please don't share it. The pages I've asked you to read lay out my (current) working practices, which we can talk about as *one example* of how one might organize one's efforts. The specifics of one's system are less important than *having* a system and following it consistently. Any thoughts you have about how the approach I describe could be improved would be very welcome.

- Additional references

The “More” section refers to works that I think you would find useful if you wanted to pursue a session’s topics at greater depth. They may or may not be available online. You are NOT responsible for reading them.

- aRrgh: a newcomer’s (angry) guide to R. <http://arrgh.tim-smith.us>

From the abstract: “R is a shockingly dreadful language for an exceptionally useful data analysis environment. The more you learn about the R language, the worse it will feel. The development environment suffers from literally decades of accretion of stupid hacks from a community containing, to a first-order approximation, zero software engineers... The goal of the document is to describe R’s data types and structures while offering enough help with the syntax to get a programmer coming from another, saner language into a more comfortable place.”

- In Our Time (2013) “The Scientific Method”. <http://www.bbc.co.uk/programmes/b01bl1jm#in=collection:p0lgyd7j>

In our time is a long-running and, I think, fantastic program on BBC Radio 4. In this episode, host Melvyn Bragg and his guests discuss the development of the scientific method. There’s a lot of good material here, but his guests do wander a bit more than they might have. You can also find it on iTunes at <https://itunes.apple.com/us/podcast/the-scientific-method/id73330895?i=353657691&mt=2>

- Long JS. 2009. The workflow of data analysis using Stata. *Stata Press books*.

Excellent on *how to think* about establishing and maintaining your data analysis projects effectively. His insights are particularly valuable for semi-perpetual projects with large research teams, so parts of it may be less valuable to those working on more discrete projects. Whatever the project, consciously developing procedures to address the concerns he raises will serve you well.

- Mitchell MN. 2010. *Data management using stata: A practical handbook*. Stata Press: College Station, Tex.

An excellent complement to Long. More oriented towards the practical mechanics of managing the data analysis process.

- Platt JR. 1964. Strong inference. *Science* **146**(3642): 347-353

One of the papers discussed in Stearns’ presentation.

- UCLA’s “Resources to help you learn and use Stata” <http://www.ats.ucla.edu/stat/stata/>

UCLA’s Institute for Digital Research and Education maintains a marvelous set of pages on many aspects of using Stata (as well as other software packages). Highly recommended.

Assignment

- Assignment 1: Pre-class evaluation

Session 2 A review of Ordinary Least Squares

Jan 16, 2017

Because Ordinary Least Squares should be familiar to all of us, it's a great setting to revisit some foundational concepts that will be with us for the entire semester.

Reading

- A review of ordinary least squares

A lot of reading, but most of it should be quite familiar. The goal is just to re-familiarize ourselves with some of the key concepts of regression, before we take off on variations to them. Rather than focusing on what is familiar, please come to class with questions about what was unfamiliar or confusing.

- Carlson KD, Wu JP. 2012. The illusion of statistical control: Control variable practice in management research. *Organizational Research Methods* **15**(3): 413-435
- Kennedy, Chapters 1 Introduction, 2 Criteria for Estimators, 3 The Classical Linear Regression Model and 4 Interval Estimation and Hypotheses Testing
- Wooldridge, Chapters 1 The Nature of Econometrics and Economic Data, 2 The Simple Regression Model, 3 Multiple Regression Analysis: Estimation and 4 Multiple Regression Analysis: Inference

- Additional references

Some useful papers on specific issues that often arise in OLS and on Stata

- Aguinis H, Gottfredson RK, Joo H. 2013. Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods* **16**(2): 270-301
- Cox NJ. 2002. Speaking Stata: How to face lists with fortitude. *Stata Journal* **2**(2): 202-222

We often want to repeat an operation in Stata, e.g., applying the same transformation to multiple variables or importing multiple files. This article introduces the commands Stata provides for doing so, along with macros, which are Stata's equivalent to "variables" in other computing languages (since "variables" means something else in the statistical context). By minimizing the need to retype commands, macros enable efficiency and flexibility, while reducing chances for error.

- Kennedy PE. 2005. Oh no! I got the wrong sign! What should I do? *Journal of Economic Education* **36**(1): 77-92

Forty explanations for why your results might not turn out as you'd expected.

- Rosopa PJ, Schaffer MM, Schröder AN. 2013. Managing heteroscedasticity in general linear models. *Psychol Methods* **18**(3): 335-351
- Silva JMCS, Tenreiro S. 2006. The log of gravity. *Review of Economics and Statistics* **88**(4): 641-658
- Spector PE, Brannick MT. 2010. Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods* **14**(2): 287-305

Assignment

- Assignment 2: Managing data in Stata

Session 3 More on OLS and current controversies

Jan 23, 2017

Reading

- Interaction terms

I know this was covered in Research Methods I, but there are a several points I want to (re)establish. I've provided the Brambor *et al* and Braumoeller papers to refresh your memory, as both are models of clear exposition. Reading just one of them would probably be sufficient. Williams 2015 is a set of lecture notes he was kind enough to share on the Internet. A quick skim is probably sufficient, but I wanted to include it because it is easy to follow and gives examples using Stata.

While the prior papers discussed (mean) centering of variables to make interpretation of interaction effects easier, several prominent authors have suggested that mean-centering variables before interacting them helps resolve the common problem of multicollinearity. Despite several papers conclusively demonstrating that it doesn't actually make any difference, the misconception persists, so it's important that you understand why it doesn't help. Echambadi and Hess provide a clear explanation.

Lastly, I've included Williams 2012 mostly for your reference. Stata's *margins* and *marginsplot* commands help in calculating and displaying the marginal effects of estimated coefficients. They are particularly useful when interactions are involved. Stata's documentation for these commands is thorough to the point of overwhelming initially. Although it was written for Stata 11, Williams' piece still provides a useful introduction to these commands. If you've at least skimmed this paper before class, you'll get more out of the in-class demonstration of these commands. But, you don't need to read it in-depth. By the way, Williams received the 2015 *Stata Journal* Editors' Prize, which he richly deserved. You could do much worse than reading pretty much everything methods-related on his website, [<https://www3.nd.edu/~rwilliam/>][19].

- Brambor T, Clark WR, Golder M. 2006. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis* 14(1): 63-82
- Braumoeller BF. 2004. Hypothesis testing and multiplicative interaction terms. *International Organization* 58(4): 807-820
- Echambadi R, Hess JD. 2007. Mean-centering does not alleviate collinearity problems in moderated multiple regression models. *Marketing Science* 26(3): 438-445

Models with interaction terms are often plagued with high multicollinearity. Several prominent authors have suggested that mean-centering the variables before interacting them helps resolve this problem. They are wrong. Despite several

papers, including this one, conclusively demonstrating that it doesn't actually make any difference, the misconception persists.

- Kromrey JD, Foster-Johnson L (1998) "Mean Centering in Moderated Multiple Regression: Much Ado about Nothing." *Educational and Psychological Measurement*. 58(1):42-67.
- Williams R. 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal* 12(2): 308-331

- Comparing coefficients across groups

I've combined this topic with our discussion of interaction terms since a common use of interactions is to determine if a covariate's effect differs across groups, e.g. men versus women or foreign versus domestic firms. These are all short papers and fairly quick reads. Order isn't significant.

- Aguinis H, Pierce CA. 1998. Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods* 1(3): 296-314

Okay, so this paper isn't particularly short or that easy a read. However, you only need to read up to "Alleviating Heterogeneity of Error Variance Effects: Solutions Recently Proposed" on page 307.

- Clogg CC, Petkova E, Haritou A. 1995. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* 100(5): 1261-1293

The only section you need to read is "B. A Note on Comparing Regression Coefficients between Groups" on pages 1276-7.

- Gelman A, Stern H. 2006. The difference between "significant" and "not significant" is not itself statistically significant. *American Statistician* 60(4): 328-331
- Wolfe R, Hanley J. 2002. If we're so different, why do we keep overlapping? When 1 plus 1 doesn't make 2. *Canadian Medical Association Journal* 166(1): 65-66

- Current controversies

"Statistically significant results" and p -values have long been at the heart of the published literature in many fields. Unfortunately, there are *many* ways they can lead us astray. Start with Gelman and Loken, who discuss the many ways even well-intentioned scholars can artificially generate "significant results". Then read Cumming. In particular, please focus on pages 11 through 14, where he demonstrates the "dance of the p -values" and proposes an "Eight-step new-statistics strategy for research with integrity." Does it seem reasonable? Likely be accepted? Next read the Bettis *et al* paper in light of all of the above (the authors are the senior editors of *SMJ*). It both elaborates on the many shortcomings and introduces a new set of policies at *SMJ*. The most striking element for most, I suspect, will be that "SMJ will no longer accept papers for publication that report or refer to cutoff levels of statistical

significance (p-values).” SMJ is not the first journal to move in this direction and I suspect other journals will follow over time, at least in part. So, understanding the rationale and implications of SMJ’s new policies is career-enhancing, even if it raises headaches in the immediate. (As an Associate Editor at SMJ, I can promise you—it causes headaches in the immediate). One effect of the SMJ policy will be, I believe, more reporting of confidence intervals. With that in mind, next read Høekstra *et al*, who discuss issues in the field’s understanding of confidence intervals. Lastly, read Gelman (2004) for another way to think about errors in our analysis.

As you read these papers, please keep in mind our discussions in the first session about theory, method and their intersection. What are we ultimately trying to accomplish with we test a hypothesis and what does that mean for the most appropriate tools to use.

- Bettis RA, Ethiraj S, Gambardella A, Helfat C, Mitchell W. 2015. Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*: n/a-n/a
 - Cumming G. 2014. The new statistics: Why and how. *Psychological Science* **25**(1): 7-29
 - Gelman A, Loken E. 2014. The statistical crisis in science. *American Scientist* **102**(6): 460
 - Gelman A. 2004. Type 1, type 2, type S, and type M errors. *Statistical Modeling, Causal Inference, and Social Science* http://andrewgelman.com/2004/12/29/type_1_type_2_t/.
 - Høekstra R, Morey RD, Rouder JN, Wagenmakers EJ. 2014. Robust misinterpretation of confidence intervals. *Psychon Bull Rev* **21**(5): 1157-1164
- Additional references
 - Bakan D. 1966. The test of significance in psychological research. *Psychological bulletin* **66**(6): 423

One of the classic pieces on the (mis)use of tests of statistical significance. The section, “Difficulties Of The Null Hypothesis” beginning on page 425 is particularly relevant.

 - Dalal DK, Zickar MJ. 2012. Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods* **15**(3): 339-362
- Another paper making essentially the same point as the Echambadi and Hess paper above.
- Falk R, Greenbaum CW. 1995. Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology* **5**(1): 75-98

Another paper on the p-values controversy. Particularly thoughtful on why the practice has persisted despite over 80 years of authors trying to explain why it shouldn't.

- Gelman A, O'Rourke K. 2014. Discussion: Difficulties in making inferences about scientific truth from distributions of published p-values. *Biostatistics* **15**(1): 18-23; discussion 39-45
- Gelman A, Weakliem D. 2009. Of beauty, sex and power. *American Scientist* **97**(4): 310-316
- Gigerenzer G. 2004. Mindless statistics. *The Journal of Socio-Economics* **33**(5): 587-606
- Goldfarb B, King AA. 2016. Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal* **37**(1): 167-176
- Hoekstra R, Morey RD, Rouder JN, Wagenmakers EJ. 2014. Robust misinterpretation of confidence intervals. *Psychon Bull Rev* **21**(5): 1157-1164

Another push for, among other things, reporting confidence intervals rather than p-values. Note, however, evidence from Hoekstra *et al* in the additional material, that researchers aren't much more sure about what CIs mean than they are of p-values.

- Hubbard R, Lindsay RM. 2008. Why p values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology* **18**(1): 69-88

Another paper on the p-values controversy. I found pages 80-81 particularly telling.

- Killeen PR. 2005. An alternative to null-hypothesis significance tests. *Psychol Sci* **16**(5): 345-353

Interesting to see that this has been cited over 200 times in psychology.

- Kline RB. 2013. *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). American Psychological Association: Washington, DC, US.

See Chapter 4 in particular, although this entire book merits very thoughtful reading.

- Krzywinski M, Altman N. 2013. Points of significance: Significance, p values and t-tests. *Nature Methods* **10**(11): 1041-1042
- National Academies of Sciences, Engineering and Medicine. 2016. *Statistical challenges in assessing and fostering the reproducibility of scientific results: Summary of a workshop*. The National Academies Press: Washington, DC

The National Academies held an interesting workshop on reproducibility. I include this just as a good reference of current thinking. I don't necessarily agree

with all of it, e.g., demanding stricter p-values sort of misses the point, but it's valuable.

- Savalei V, Dunn E. 2015. Is the call to abandon p-values the red herring of the replicability crisis? *Front Psychol* **6**: 245
- Williams R. 2015. Interpreting interaction effects; interaction effects and centering. <https://www3.nd.edu/~rwilliam/stats2/l53.pdf>

Notes that Williams prepared for his class and was kind enough to post on the Internet. Focus primarily on his discussion of centering. Hopefully, this will mostly be familiar and thus a quick read, but I find his presentation wonderfully clear.

Assignment

- Assignment 3: Interpreting OLS

Bayesian analysis

Session 4 A brief look at Bayesian analysis

Jan 30, 2017

Almost all management research to date has been based on *frequentist* principals. *Bayesian* modeling offers an alternative approach. While we'll focus overwhelmingly on frequentist approaches in this class, I wanted to give you an introduction to Bayes models for several reasons. First, they offer a coherent solution to the problems in null hypothesis testing that we discussed last time—indeed, several of last session's articles include Bayesian methods in their proposals for improved practice. Second, many barriers to their use have fallen. The quasi-philosophical debate about them has mostly settled and the software available to run them is more accessible (Stata 14 even introduced some Bayesian functionality. Rudimentary at the moment, but I'm sure it will improve). Third, they allow researchers to do things that are difficult or impossible to do with frequentist techniques, e.g., confidently accepting a null hypothesis and imposing inequality constraints. Fourth, its use is growing rapidly in other fields. Political science and ecology are ahead of management in its use and it is key to many of the machine learning/predictive analysis techniques being driven by the availability of big data. Lastly, there are stirrings of interest in the management literature. For example, several of today's readings come from 2014 *Journal of Management* special Issue on Bayesian methodology (of course, the issue includes a cautionary editorial on why they might not make inroads into the literature).

Unlike our other sessions, the goal for today is not to equip you to do a Bayesian analysis. It's not particularly difficult, but a sufficient grounded is more than a day's work. Rather, I want to give you a basic understanding of how Bayes analysis differs from frequentist technique, the advantages it offers and some pointers for further investigation if you are interested.

Reading

To set the stage, read “Dr. Fisher's casebook” by Anonymous (actually, the editors of the journal *Significance*). Then move to Zyphur and Oswald, who give a nice overview of Bayesian analysis vis-a-vis the more familiar frequentist approach. Then read Skorupski & Wainer and Andraszewicz *et al* for an of Bayes factors. Then read Kruschke's piece on how to credibly *accept* the null hypothesis. Lastly, skim McKee and Miller on some of the institutional opportunities and challenges facing Bayesian analysis in management.

- Method
 - Andraszewicz S, Scheibehenne B, Rieskamp J, Grasman R, Verhagen J, Wagenmakers EJ. 2014. An introduction to Bayesian hypothesis testing for management research. *Journal of Management* **41**(2): 521-543

- Anonymous. 2008. Dr Fisher's casebook: The doctor minds his ps. *Significance* 5(3): 126-126
- Etz A, Gronau QF, Dablander F, Edelbrunner PA, Baribault B. 2016. How to become a Bayesian in eight easy steps: An annotated reading list.
- Kruschke JK, Liddell TM. Forthcoming. The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin and Review*
- Kruschke JK. 2011. Bayesian assessment of null values via parameter estimation and model comparison. *Perspect Psychol Sci* 6(3): 299-312
- McKee RA, Miller CC. 2014. Institutionalizing Bayesianism within the organizational sciences: A practical guide featuring comments from eminent scholars. *Journal of Management* 41(2): 471-490
- Skorupski WP, Wainer H. 2015. The Bayesian flip: Correcting the prosecutor's fallacy. *Significance* 12(4): 16-20
- Wagenmakers E-J. 2007. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review* 14(5): 779-804
- Zyphur MJ, Oswald FL. 2013. Bayesian estimation and inference: A user's guide. *Journal of Management*
- Use
 - Hansen MH, Perry LT, Reese CS. 2004. A Bayesian operationalization of the resource-based view. *Strategic Management Journal* 25(13): 1279-1295
- Additional references

I include two beginner-friendly textbooks, along with some articles of interest. Given the massive advances in Bayesian analysis over the last decade, avoiding older references may be avoidable, at least until you are on confident enough footing to know what's been superseded.

- Gelman A. 2014. *Bayesian data analysis* (Third edition. ed.). CRC Press: Boca Raton
 The second textbook I'd use. A bit less intuitive than Kruschke, but still pretty accessible.
- Gross J. 2013. A unified test for substantive and statistical significance in political science. Working Paper.
- Hahn ED, Doh JP. 2006. Using Bayesian methods in strategy research: An extension of hansen et al. *Strategic Management Journal* 27(8): 783-798
- Jebb AT, Woo SE. 2014. A Bayesian primer for the organizational sciences: The "two sources" and an introduction to bugsxla. *Organizational Research Methods* 18(1): 92-132

- Kruschke JK. 2012. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*
- Kruschke, John K. 2014. *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*. Elsevier

This is the textbook I would start with. Accessible and rigorous.

Assignment

- Assignment 4: A s(t)imulating experience

Repeat appearances: Multilevel and panel data

In this module, we'll discuss circumstances in which we have multiple occurrences of related or identical observations. Two previously distinct literatures in this areas have actually moved closer together over time because they were often solving the same problems from only slightly different perspectives. However, they remain distinct enough that we'll discuss them separately while drawing the relevant connections between them.

Multilevel models apply when observations are “nested” within larger units. For example, much of the pioneering work in this area was done in education, since all students in a given classroom experience a common effect from observed and unobserved factors that differ between classrooms (think of class size and the underlying emotional tenor of the class, respectively). At a minimum, we need to control for this. Beyond that, we may be explicitly interested in cross-level effects, e.g., does a student's socio-economic status matter more as classroom size increases?

The nesting can also represent multiple observations of the same unit. Think of studying a group of firms over a ten year period. You can think of ten annual observations being “nested” under a given firm and thus potentially influenced by unobserved characteristics of that firm. This is the main point of intersection between multilevel models—which, as a massive over-generalization, grew largely out of psychology and education—and *panel data*, which developed primarily out of economics—again, as a massive over-generalization.

Of course, all of the observations made in a given year are also “nested” within that year and may be influenced by unobserved characteristics of that year. Furthermore, there may be temporal spillovers across a firm's observations. Above and beyond what is common to all observations of Firm X, unobservable factors that affected Firm X's profitability in year Y might persist into year $Y+1$.

Broadly speaking, it is this emphasis on the temporal component that characterizes panel data, while greater emphasis on cross-level effects characterizes multilevel modeling. However, as the approaches continue to cross-pollinate, the lines between them become increasingly blurred.

While panel data and multilevel methods allow us to understand complex, dynamic causal relationships, an initially separate literature developed for the much simpler purpose of generating correct standard errors despite the presence of non-independent observations. This approach operates by “clustering” the standard errors across related observations. There are two main points of contact between this literature and panel data/multilevel methods. First, which technique should we use? The question “Should I cluster standard errors or used a random effects model?” may be the most common question on StataList. Second, as the literatures had become more intertwined over time, the question of whether they should be used in combination has surfaced. We'll spend some time on the first question and I'll point you to literature on the second.

Session 5 Panel data 1

Feb 6, 2017

Reading

We'll cover panel data and clustered standard errors today. Regarding panel data, start with Frees, who will give you a pretty complete introduction to key issues. Skim through Kennedy and Wooldridge (mostly chapter 14 — I'm not sure why he presented the material in chapter 13 when & how he did). Read Baum mostly for Stata implementation details. Then read Certo & Semadeni for implications for strategy. Regarding clustering standard errors, read Arceneaux & Nickerson and Cameron & Miller in whichever order you prefer.

- Method

- Arceneaux K, Nickerson DW. 2009. Modeling certainty with clustered data: A comparison of methods. *Political Analysis* 17(2): 177-190
- Cameron CA, Miller DL. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* 50(2): 317-372
- Certo ST, Semadeni M. 2006. Strategy research and panel data: Evidence and implications. *Journal of Management* 32(3): 449-471
- Frees EW. 2004. *Longitudinal and panel data : Analysis and applications in the social sciences*. Cambridge University Press: West Nyack, NY, USA. Chapters 1-3. <http://site.ebrary.com/lib/asulib/docDetail.action?docID=10131690>
- Kennedy, Chapter 18 Panel Data
- Wooldridge, Ch 13 Pooling Cross Sections across Time: Simple Panel Data Methods and 14 Advanced Panel Data Methods

Odd order of presentation. May do better with 14 first (and perhaps only) after reading Kennedy and Baum.

- Use

- O'Brien JP, David P. 2014. Reciprocity and r&d search: Applying the behavioral theory of the firm to a communitarian context. *Strategic Management Journal* 35(4): 550-565
- Stadler C, Helfat CE, Verona G. 2013. The impact of dynamic capabilities on resource access and development. *Organization Science* 24(6): 1782-1804

- Additional references

Two additional descriptions of panel data models in general. The Petersen is nice for its focus on different ways of handling the error term, particularly possible temporal effects. The Greene paper addresses an issue that doesn't get enough attention, how

to handle fixed and random effects in non-linear models. We'll address many such models later in the semester, including logit, probit, etc. I also include a nice review of the commands available for panel data (at least fixed effects) in Stata for your reference.

- Halaby CN. 2004. Panel models in sociological research: Theory into practice. *Annual Review of Sociology* **30**(1): 507-544
- Petersen MA. 2008. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* **22**(1): 435-480
- McCaffrey DF, Lockwood JR, Mihaly K, Sass TR. 2012. A review of Stata commands for fixed-effects estimation in normal linear models. *Stata Journal* **12**(3): 406-432
- Greene W. 2001. Fixed and random effects in nonlinear models. Stern School of Business, New York University

Assignment

- Assignment 5: Wrap up

Session 6 Panel data 2 and Multilevel Models 1

Feb 13, 2017

Reading

Today we'll complete our discussion of panel data models and introduce multilevel models. Please start by reading the two pieces under Method (panel data). Both papers elaborate on the issue of clustering versus fixed effects versus random effects. Then read the two Use (Panel data) papers. To introduce the basic concepts of multilevel models, then read Hofmann.

- Method (Panel data)
 - Bartels BL. 1998. Beyond “fixed versus random effects”: A framework for improving substantive and statistical analysis of panel, time-series cross-sectional, and multilevel data. Working paper.
 - Bryan ML, Jenkins SP. 2015. Multilevel modelling of country effects: A cautionary tale. *European Sociological Review*
 - Schunck R, Perales F, others (2017) “Within-and between-cluster effects in generalized linear mixed models: A discussion of approaches and the xthybrid command.” *Stata Journal*. 17(1):89-115.
- Use (Panel data)
 - Fabrizio KR. 2012. Institutions, capabilities, and contracts: Make or buy in the electric utility industry. *Organization Science* 23(5): 1264-1281
 - Kim C, Bettis RA. 2014. Cash is surprisingly valuable as a strategic asset. *Strategic Management Journal* 35(13): 2053-2063
- Additional references (Panel data)
 - Achen CH. 2000. Why lagged dependent variables can suppress the explanatory power of other independent variables. Working paper

A recurrent issue in panel models is when and if it is appropriate to include lagged values of the dependent variable as a regressor. This working paper, which has been cited almost 700 times, but never published(?!), lays out the issues pretty clearly. If you did into subsequent literature on this issue, you'll frequently see it cited. Achen is an extremely well respected methodologist in political science.
 - Ballinger GA. 2004. Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods* 7(2): 127-150
 - Beck N, Katz JN. 1995. What to do (and not to do) with time-series cross-section data. *American Political Science Review* 89(3): 634-647

A classic reference. The Certo *et al* paper we read, among others, has raised questions about the specific recommendations of this paper, but you need to be aware of it.

- Bell A, Jones K. 2015. Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods* 3(01): 133-153
- Burdisso T, Sangiácomo M. 2016. Panel time series: Review of the methodological evolution. *Stata Journal* 16(2): 424-442
- Schunck R. 2013. Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *Stata Journal* 13(1): 65-76

Interest in Mundlak (1978, *Econometrica*) model of correlated random effect has resurged recently. The model offers a compromise of sorts between fixed and random effects. We won't have time to discuss in detail, so I include this paper as a recent introduction and implementation in Stata.

- Method (Multilevel)

- Hofmann DA. 1997. An overview of the logic and rationale of hierarchical linear models. *Journal of Management* 23(6): 723-744

- Additional references (Multilevel)

- Aguinis, H., & Molina-Azorin, J. F. 2015. Using multilevel modeling and mixed methods to make theoretical progress in microfoundations for strategy research. *Strategic Organization*, 13: 353-364.
- Aguinis, H., & Culpepper, S.A. 2015. An expanded decision making procedure for examining cross-level interaction effects with multilevel modeling. *Organizational Research Methods*, 18: 155-176.
- Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. 2013. Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management*, 39: 1490-1528.
- Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. 2009. First decade of Organizational Research Methods: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods*, 12: 69-112.
- Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen. G. 2012. Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, 97: 951-966.
- Morgeson, F. P., & Hoffman, D. A. 1999. The structure and function of collective constructs: Implications for multilevel research and theory development. *Academy of Management Review*, 24: 249-265.

- Skrondal A, Rabe-Hesketh S. 2008. Multilevel and related models for longitudinal data. In J de Leeuw, E Meijer (Eds.), *Handbook of multilevel analysis: 275-299*. Springer <http://site.ebrary.com/lib/asulib/docDetail.action?docID=10222955>

Assignment

- Assignment 6: Panel data 1

Session 7 Multilevel 2

Feb 20, 2017

Reading

Please start with Frees, which you should find a slightly more formal/technical presentation of the material in the Hofman article we read last week. Then read Enders and Tofighi, who discuss centering, which ends up being very important in multilevel models. Next, read Bliese and Ployhart, who discuss the modeling of growth rates over time. Lastly, *skim* McNeish *et al*, who argue that HLM is often used reflexively in situations where it is *not* in fact the optimal modeling approach. At this point, awareness and a *basic* understanding of HLM's limitations is sufficient.

- Method

- Bliese PD, Ployhart RE. 2002. Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods* 5(4): 362-387
- Enders CK, Tofighi D. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychol Methods* 12(2): 121-138
- Frees EW. 2004. *Longitudinal and panel data : Analysis and applications in the social sciences*. Cambridge University Press: West Nyack, NY, USA. Ch 5. <http://site.ebrary.com/lib/asulib/docDetail.action?docID=10131690>
- McNeish D, Stapleton LM, Silverman RD. 2016. On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*

- Use

- Anderson SW, Glenn D, Sedatole KL. 2000. Sourcing parts of complex products: Evidence on transactions costs, high-powered incentives and ex-post opportunism. *Accounting, Organizations and Society* 25(8): 723-749
- Gielnik MM, Barabas S, Frese M, Namatovu-Dawa R, Scholz FA, Metzger JR, Walter T. 2014. A temporal analysis of how entrepreneurial goal intentions, positive fantasies, and action planning affect starting a new venture and when the effects wear off. *Journal of Business Venturing* 29(6): 755-772
- Kumar MVS. 2013. The costs of related diversification: The impact of the core business on the productivity of related segments. *Organization Science* 24(6): 1827-1846

- Additional references

- Bollen KA, Brand JE. 2010. A general panel model with random and fixed effects: A structural equations approach. *Social Forces* 89(1): 1-34

- Cudeck R, Harring JR. 2007. Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology* **58**: 615-637
- Culpepper SA, Aguinis H. 2016. Should I use multilevel modeling? A hands-on demonstration.
- Panik MJ. 2013. *Growth curve modeling : Theory and applications*. John Wiley & Sons, Incorporated: Somerset, NJ, USA

An accessible reference if you need to explore growth models in more detail.

- Singer, J. D. 1998. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, **24**: 323-355.
- Stegmueller D. 2013. How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science* **57**(3): 748-761

Assignment

- Assignment 7: Panel data 2

Limited dependent variables

OLS assumes that the dependent variable can take any real value. In this section of the class, we will examine models for dependent variables that are “limited” in the values they can take. For example, they may be a probability, which is bounded between 0 and 1, or a count, which can only be a non-negative integer.

Session 8 Logit and probit – basics and interpretation

Feb 27, 2017

When the dependent variable is the probability of an outcome occurring, the most commonly used models are the logit and probit. While not intrinsically hard, they do have certain nuances that you need to understand to use them properly. We'll spend this session and the next working through these.

Reading

Start with Long, who gives you almost all of what you need. Having read Long, you will probably find Bowen & Wiersema (read through page 103) pretty familiar. But, if you are confused by anything in Long, this is a very clearly written practical guide that should help clarify things for you. Wooldridge and Kennedy collectively provide useful background and additional detail. Pinzon is a short, practical guide to interpretation via Stata. Close your reading on the logit model with Høetker 2007. I've also included the Myung paper because logit and probit are the first models we've encountered that are usually estimated via Maximum Likelihood.

- Method
 - Bowen HP, Wiersema MF. 2004. Modeling limited dependent variables: Methods and guidelines for researchers in strategic management. *Research Methodology in Strategy and Management* **1**: 87-134
 - Høetker G. 2007. The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal* **28**(4): 331-343
 - I've actually included this paper multiple times in the readings for this class and the next, because it references multiple issues in the use of logit/probit models. For today, please skip the sections on interaction terms and comparing coefficients across groups. We'll discuss those next time.
 - Kennedy, Chapter 16. Qualitative Dependent Variables, section 16.1.
 - Long, Chapters 4 5 and 6
 - Myung, I. J. 2003. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47: 90-100.
 - Pinzon E. 2016. Effects of nonlinear models with interactions of discrete and continuous variables: Estimating, graphing, and interpreting. *The Stata Blog*. <http://blog.stata.com/2016/07/12/effects-for-nonlinear-models-with-interactions-of-discrete-and-continuous-variables-estimating-graphing-and-interpreting/>
 - Wooldridge, Chapter 17 Limited Dependent Variable Models and Sample Selection Corrections, section 17.1

- Use
 - Gulati R. 1995. Does familiarity breed trust - the implications of repeated ties for contractual choice in alliances. *Academy of Management Journal* **38**(1): 85-112
 - Marcel JJ, Cowen AP. 2014. Cleaning house or jumping ship? Understanding board upheaval following financial fraud. *Strategic Management Journal* **35**(6): 926-937
- Additional references
 - Agresti A. 2014. *Wiley series in probability and statistics : Categorical data analysis (3rd edition)*. Wiley: Somerset, NJ, USA. <http://site.ebrary.com/lib/asulib/docDetail.action?docID=10867993>
A thorough textbook introduction.
 - Breen R, Karlson KB, Holm A. 2013. Total, direct, and indirect effects in logit and probit models. *Sociological Methods and Research* **42**(2): 164-191
Mediation can get odd when limited dependent variables are involved. This is a useful paper on how to address it.
 - Buis ML. 2010. Direct and indirect effects in a logit model. *Stata Journal* **10**(1): 11-29
Mediation can get odd when limited dependent variables are involved. This is a useful paper on how to address it
 - Hellevik O. 2007. Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity* **43**(1): 59-74
There are times when estimating a probability with OLS, the so-called linear probability model, is attractive. This paper discusses the pluses and minuses of doing so.
 - Hilbe J. 2009. *Logistic regression models*
Somewhat more technical, but still accessible. Amazingly thorough.
 - Hosmer DW, Lemeshow S, Sturdivant RX. 2013. *Wiley series in probability and statistics : Applied logistic regression (3rd edition)*. John Wiley & Sons: New York, NY, USA <http://site.ebrary.com/lib/asulib/docDetail.action?docID=10677827>
An absolute classic. These are among the clearest and most authoritative authors out there.
 - Huang C, Shields TG. 2000. Interpretation of interaction effects in logit and probit analyses. Reconsidering the relationship between [us] registration laws, education, and voter turnout. *American Politics Quarterly* **28**(1): 72-79
An excellent resource for examples on how to calculate and graphically present the effect of an interaction term in logic & probit models

- MacKinnon DP, Lockwood CM, Brown CH, Wang W, Hoffman JM. 2007. The intermediate endpoint effect in logistic and probit regression. *Clinical Trials* 4(5): 499-513

Mediation can get odd when limited dependent variables are involved. This is a useful paper on how to address it.

- McFadden DL, Manski CF. 1981. *Structural analysis of discrete data and econometric applications*

Slightly advanced, but you could do worse than learn from someone (McFadden) who received the Nobel prize in economics for his work on the subject. Especially good for the coverage of Independence of Irrelevant Alternatives (IIA). McFadden was kind enough to make this book freely available for educational purposes.

- Menard S. *Logistic regression: From introductory to advanced concepts and applications*. *Logistic regression: From introductory to advanced concepts and applications*. Sage publications, inc. SAGE Publications, Inc.: Thousand Oaks, CA

Another good textbook choice. Available from online via ASU Libraries [[*http://dx.doi.org/10.4135/9781483348964](http://dx.doi.org/10.4135/9781483348964)][113].*

- Mood C. 2010. Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review* 26(1): 67-82

Mood focuses heavily on the underlying issue of *unobserved heterogeneity*, so if you are struggling with what that is and why it matters, you should find it especially useful.

- Train K. 1986. *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand*. MIT Press: Cambridge, MA

Train is an extremely clear writer. As a transportation economist, he comes at things a little bit differently than many other presentations, which I find helpful.

- Train K. 2009. *Discrete choice methods with simulation*. Cambridge university press

Train is an extremely clear writer. As a transportation economist, he comes at things a little bit differently than many other presentations, which I find helpful.

Assignment

- Assignment 8: TBD

Spring Break

Mar 6, 2017

Session 9 Logit and probit – complications and extensions

Mar 13, 2017

Reading

This looks like a tremendous amount of reading, but for most of the entries, I'm only asking you to read a short section.

- Interaction effects in logit models

I would read these papers in alphabetical order.

- Hoetker G. 2007. The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal* **28**(4): 331-343

You read most of this paper last week. For here, please just read the section on interaction effects.

- Norton EC, Wang H, Ai C. 2004. Computing interaction effects and standards errors in logit and probit models. *Stata Journal* 4(2): 154-167

I include this paper because it is often referenced, but—frankly—is somewhere between wrong and incomplete. Skim it enough to get the basic idea and we'll discuss its limitations in class.

- Puhani PA. 2012. The treatment effect, the cross difference, and the interaction term in nonlinear “difference-in-differences” models. *Economics Letters* **115**(1): 85-87

One of several prominent papers criticizing Norton *et al.* I include it mostly so you are familiar with it when a reviewer or co-author suggests you use Ai and Norton's approach. Skim it and we'll discuss in detail in class.

- Comparing coefficients across groups

I would start with the relevant section of Hoetker 2007, which introduces the problem and several possible solutions. Then move on to Long, Williams and Zelner, who each offer different solutions.

- Hoetker G. 2007. The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal* **28**(4): 331-343

Yes, this paper again. Please just read the relevant section.

- Long J. 2009. Group comparisons in logit and probit using predicted probabilities. Working paper draft 2009-06-25. Retrieved on September 15, 2009
- Williams R. 2009. Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups. *Sociological Methods & Research* 37(4): 531-559

- Zelner BA. 2009. Using simulation to interpret results from logit, probit, and other nonlinear models. *Strategic Management Journal* **30**(12): 1335-1348
- Extending to more than 2 choices

Until now, we've focused on situations where there are two outcomes. With some modification, logit and probit models can also address cases in which there are more than two choices. Start with Long, chapter 7, in which he discusses ordinal outcomes, meaning those with some natural ordering such as poor, average, or good. Then read Long, chapter 8, in which he covers nominal outcomes, that is, those without any natural ordering such as investing in Germany, France or Italy. The relevant sections of Bowen & Wiersema (pages 103-119) should, again, help clarify anything left unclear by Long. Kennedy and Wooldridge will provide additional explanations on both ordinal and nominal outcomes. Skim the Stata manual entries for *mlogit* and *ologit*.

 - Bowen HP, Wiersema MF. 2004. Modeling limited dependent variables: Methods and guidelines for researchers in strategic management. *Research Methodology in Strategy and Management* **1**: 87-134
 - Kennedy, Ch. 16, sections 16.2 and 16.3
 - Stata Manual [R] mlogit
 - Stata Manual [R] ologit
 - Wooldridge, Ch 17, Limited Dependent Variable Models and Sample Selection Corrections, Section 17.1
- Use
 - Dushnitsky G, Shapira Z. 2010. Entrepreneurial finance meets organizational reality: Comparing investment practices and performance of corporate and independent venture capitalists. *Strategic Management Journal* **31**(9): 990-1017
An application of ordered logit.
 - Hoetker G. 2005. How much you know versus how well I know you: Selecting a supplier for a technically innovative component. *Strategic Management Journal* **26**(1): 75-96
An application of comparing coefficients.
 - Parmigiani A. 2007. Why do firms both make and buy? An investigation of concurrent sourcing. *Strategic Management Journal* **28**(3): 285-311
An example of multinomial logit.
- Additional references
 - Amemiya T. 1981. Qualitative response models: A survey. *Journal of Economic Literature* **19**: 1483-1536

Old, but sweeping review of the range of models related to those we've been discussing. Amemiya is also one an authoritative source.

- Greene WH, Hensher DA. 2008. Modeling ordered choices: A primer and recent developments. SSRN. <http://ssrn.com/paper=1213093>.
- Jackman S. 2007. Models for unordered outcomes. <http://jackman.stanford.edu/classes/350C/07/unordered.pdf>

I simply haven't found a much better presentation of this topic than this. These are notes that Professor Jackman prepared for his class and was kind enough to post on the Internet.

- Liu X. 2016. *Applied ordinal logistic regression using stata : From single-level to multilevel modeling*. Sage: Thousand Oaks, California

I've not read this one yet, but if you need to do ordinal logistic modeling in Stata this is—if nothing else—the most specialized text you'll find for the purpose.

- Mallick D. 2009. Marginal and interaction effects in ordered response models.
- Powers EA. 2005. Interpreting logit regressions with interaction terms: An application to the management turnover literature. *Journal of Corporate Finance* 11(3): 504-522
- Powers EA. 2005. Interpreting logit regressions with interaction terms: An application to the management turnover literature. *Journal of Corporate Finance* 11(3): 504-522
- Taplin RH. 2015. Research involving limited dependent variables: Issues in the literature and recommendations for improvement. *Quality & Quantity* 50(5): 2121-2140
- Williams R. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal* 6(1): 58-82
- Williams R. 2010. Fitting heterogeneous choice models with oglm. *The Stata Journal* 10(4): 540-567
- Winship C, Mare RD. 1984. Regression models with ordinal variables. *American Sociological Review* 49(4): 512-525

Assignment

- Assignment 9: Logit and probit 1

Session 10 Hazard rate models 1

Mar 20, 2017

We are often interested in what determines the probability of an event as a function of time—how long it will be until event occurs. Frequent examples include firm failure, time between innovations, etc. Doing so introduces all sorts of challenges, including the underlying impact of time (e.g, firm failure probably peaks when firms are “adolescents”), observations leaving the sample before experiencing the event, observations not experiencing the event before the end of your study, and observations experiencing the event multiple times. The statistical framework that has developed in response is not particularly difficult, but you may find it quite foreign. In particular, it has a distinct vocabulary of hazard rates, left- and right-censoring, survival functions, etc. We’ll use this class session to explore the conceptual and statistical foundations of survival analysis. In our next session, we’ll look at the practicalities of modeling survival data in Stata, since it differs significantly from most forms of regression.

Many different fields are interested in modeling survival (or, equivalently, failure) and the current state of the art reflects an amalgamation of their contributions and traditions. Accordingly, you’ll see this type of analysis referred to as event history, transition data analysis, failure analysis, duration analysis, survival analysis and other terms I’m sure I’ve missed. It is *not* the same as an “Event study” analysis, which is a technique developed largely in finance to study the impact of an event on the market value of a firm.

Reading

I’ve assigned three chapters of Box-Steffenmeier and Jones, which I find among the most accessible of the standard texts. Even more than usually, I encourage you to explore the additional materials listed below if there are concepts that aren’t clear. Different presentations of the same material can sometimes be really helpful. Many of the texts I list are available online. Kleinbaum & Klein and Vittinghoff (Chapter 6) are particularly good for introducing the underlying concepts.

- Method
 - Box Steffensmeier JM, Jones BS. 2004. *Event history modeling : A guide for social scientists*. Cambridge University Press: West Nyack, NY, Chapters 2-4 <http://site.ebrary.com/lib/asulib/docDetail.action?docID=10130467>
- Use
 - Bercovitz J, Mitchell W. 2007. When is more better? The impact of business scale and scope on long-term business survival, while controlling for profitability. *Strategic Management Journal* **28**(1): 61-79
 - Hoetker G, Swaminathan A, Mitchell W. 2007. Modularity and the impact of buyer-supplier relationships on the survival of suppliers. *Management Science* **53**(2): 178-191

- Mitchell W. 1989. Whether and when? Probability and timing of incumbents' entry into emerging industrial subfields. *Administrative Science Quarterly* **34**: 208-230

- Additional references

- Aalen OO, Borgan Ø, Gjessing HkK. 2008. *Survival and event history analysis : A process point of view*. Springer: New York, NY, USA <http://site.ebrary.com/lib/asulib/docDetail.action?docID=10251977>

More technical, but still relatively accessible. Good for when you need to go beyond the basic models. By the way, this would be the Aalen of the Nelson Aalen estimator, which is really *the* modern standard for non-parametrically estimating a cumulative hazard rate.

- Klein JP, Moeschberger ML. 2003. *Survival analysis : Techniques for censored and truncated data* (2nd ed.). Springer: New York <http://site.ebrary.com/lib/asulib/detail.action?docID=10251977>

- Kleinbaum DG, Klein M. 2005. *Survival analysis: A self-learning text*. Springer New York: New York, NY <http://link.springer.com/book/10.1007%2F0-387-29150-4>

I only found out about this book while I was preparing to teach this class. I'm actually quite impressed. The author's goal of creating a "self-teaching" resource means, on the one hand, that the explanations are very thorough but, on the other hand, that it takes long time to move through. However, I would keep this available as both a learning tool and reference.

- Liu X. 2012. *Survival analysis: Models and applications*. John Wiley & Sons: Chichester, West Sussex (<http://www.myilibrary.com/?ID=367641>)

- Vittinghoff E. 2012. *Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models*. Springer: New York. <http://www.myilibrary.com/?ID=31227>

I found this book to be excellent overall and not overly embedded in the biostatistical context. The chapter on survival models is worth a read if there is anything you aren't clear about.

Assignment

- Assignment 10: Logit and probit complications and extensions

Session 11 Hazard rate models 2

Mar 27, 2017

Readings

Today's readings are background for the practical exercises we'll do in class. *stset* is the command used prepare data for analysis, *stcox* unsurprisingly for Cox proportional hazards analysis and *streg* for parametric estimation.

- Method
 - Stata manual [ST] *stset*, *stcox*, *stcox PHassumption tests*, *streg*, *streg postestimation*
- Use
 - McCann BT, Vroom G. 2014. Competitive behavior and nonfinancial objectives: Entry, exit, and pricing decisions in closely held firms. *Organization Science* 25(4): 969-990
 - Nadolska A, Barkema HG. 2014. Good learners: How top management teams affect the success and frequency of acquisitions. *Strategic Management Journal* 35(10): 1483-1507

This paper has both survival and count data analysis. Focus on the survival analysis today and we'll look at the count data in that session.

- Additional references

In addition to two wonderful books on survival analysis specifically in Stata, I've included PDFs from my lectures from a previous class, in which we dedicated four sessions to survival analysis. Some of it will appear in this class, but I thought the remaining material might be of interest to some.

- 10 Hazard introduction v2.pdf
- 11 hazard non-parametric.key
- 12 hazard semi-parametric v2.pdf
- 13 hazard parametric.pdf
- Blossfeld H-P, Golsch K, Rohwer G. 2012. *Event history analysis with Stata*. Psychology Press

The authors are sociologists who work largely with life-course data. Blossfeld and Rohwer authored the Transition Data Analysis (TDA) software package, which is still the most flexible and powerful program that I'm aware of for this type of analysis. Note, however, that they focused on Stata when they updated their classic book, *Event History Analysis*.

- Cleves MA, Gould W, Gutierrez R, Marchenko Y. 2010. *An introduction to survival analysis using Stata* (3rd ed.)

The authors all work for Stata, so you can't get much more inside information. As one would expect, it is full of information on how to get the most out of Stata when doing survival analysis. Beyond that, though, it does an excellent job of explaining the theory and practicalities behind the analysis. If I were teaching an entire class on survival analysis, this is the text I would use.

- Hess W, Persson M. 2011. The duration of trade revisited. *Empirical Economics* **43**(3): 1083-1107

We probably won't have time to discuss discrete time models at much length. This paper provides a good introduction, while contrasting them to continuous models (at least Cox PH models).

Assignment

- Assignment II: Hazard rate models I

Session 12 Count data models

Apr 3, 2017

Readings

Start with Long and Cameron & Trivedi for background. The read Blevins *et al.* Kennedy and Wooldridge will help complete the picture.

- Methodology
 - Blevins DP, Tsang EWK, Spain SM. 2014. Count-based research in management: Suggestions for improvement. *Organizational Research Methods* **18**(1): 47-69
 - Cameron AC, Trivedi PK. 1998. *Regression analysis of count data*. Cambridge University Press: West Nyack, NY, USA. Chapter 3. <http://site.ebrary.com/lib/asulib/docDetail.action?docID=10073578>

There is also a 2013 edition, but it is not available online. For our purposes, the 1998 edition will do fine.
 - Kennedy, Chapter 16, “Qualitative Dependent Variables”, Section 16.4
 - Long, Chapter 9 Models for count outcomes
 - Wooldridge, Chapter 17 Limited Dependent Variable Models and Sample Selection Corrections, Section 17.3
- Use
 - Haunschild PR, Polidoro F, Chandler D. 2015. Organizational oscillation between learning and forgetting: The dual role of serious errors. *Organization Science* **26**(6)
 - Joseph J, Gaba V. 2015. The fog of feedback: Ambiguity and firm responses to multiple aspiration levels. *Strategic Management Journal* **36**(13): 1960-1978
 - Nadolska A, Barkema HG. 2014. Good learners: How top management teams affect the success and frequency of acquisitions. *Strategic Management Journal* **35**(10): 1483-1507.
- Additional references
 - Hilbe JM, MyLibrary. 2011. *Negative binomial regression*. Cambridge University Press: Cambridge, UK;New York;

Assignment

- Assignment 12: Hazard rate models 2

Failure of exogeneity

In this module we'll discuss three topics that are separate, yet related. All relate in some way to violations in the assumptions regarding randomness in our sample.

Session 13 Truncation and Censoring

Apr 10, 2017

Truncation and censoring occur when we are non-randomly lacking data *based on the data of the dependent variable*. ** They are often confused with each other, but are not the same. Adding to the fun, they are often jumbled together with Tobit models, which some methodologists (e.g., Wooldridge) feel should only be used for true corner solutions. That is, if you were studying how much people spent weekly on alcohol, you would get a cluster of observations with y equal to zero because some people don't drink. To make it even *more* fun, the term "sample selection" is often used to encompass truncation, censoring, corner solutions *and* a totally different type of selection that is not related to the dependent variable. We'll study that in detail next week, but as a preview, consider a study of how the stock market responds to acquisition decisions by a firm. We can only observe that for firms that announce the intention to acquire another firm, but it seems likely that the decision to acquire is not random. If not, our estimation of the effect of acquisition may be biased.

Reading

I would start with Winship and Mare. Note that, in line with the discussion above, they discuss more than truncation and censoring. For today, don't worry about the discussion of treatment effects and endogenous switching regressions. I've included the Berk paper mostly because of the illustrations he provides. In content, it overlaps largely with Winship and Mare. The Canette blog post explains truncation and censoring in the Stata context. Then read the relevant section of the Bowen & Wiersema paper, which starts on page 119. I would then turn to Kennedy and Wooldridge, with the same caveat as with Winship and Mare. Lastly, read McDonald and Moffitt's discussion of Tobit models.

- Methodology
 - Berk RA. 1983. An introduction to sample selection bias in sociological data. *American Sociological Review* **48**(3): 386-398
 - Bowen HP, Wiersema MF. 2004. Modeling limited dependent variables: Methods and guidelines for researchers in strategic management. *Research Methodology in Strategy and Management* **1**: 87-134
 - Canette I. 2016. Understanding truncation and censoring. *The Stata Blog*. <http://blog.stata.com/2016/12/13/understanding-truncation-and-censoring/>. Accessed: 19 Dec 2016.
 - Kennedy, Chapter 17, Limited Dependent Variables
 - McDonald JF, Moffitt RA. 1980. The uses of tobit analysis. *The Review of Economics and Statistics* **62**(2): 318

- Winship C, Mare RD. 1992. Models for sample selection bias. *Annual Review of Sociology* **18**: 327-350
- Wooldridge, Chapter 17 Limited Dependent Variable Models and Sample Selection Corrections, sections 17.2 - 17.4
- Use
 - Roper S, Hewitt-Dundas N. 2015. Knowledge stocks, knowledge flows and innovation: Evidence from matched patents and innovation panel data. *Research Policy* **44**(7): 1327-1340
 - Arora A, Fosfuri A, Gambardella A. 2001. Specialized technology suppliers, international spillovers and investment: Evidence from the chemical industry. *Journal of Development Economics* **65**(1): 31-54
 - Fehr E, List JA. 2004. The hidden costs and returns of incentives-trust and trustworthiness among ceos. *Journal of the European Economic Association* **2**(5)
 - Weigelt C, Miller DJ. 2013. Implications of internal organization structure for firm boundaries. *Strategic Management Journal* **34**(12): 1411-1434
 - Leiponen A, Helfat CE. 2010. Innovation objectives, knowledge sources, and the benefits of breadth. *Strategic Management Journal* **31**(2): 224-236
- Additional references
 - Rigobon R, Stoker TM. 2007. Estimation with censored regressors: Basic issues. *International Economic Review* **48**(4): 1441-1467
 - Xe DV. 2016. Tobit models. <http://ctu.edu.vn/~dvxe/econometrics/tobitl.pdf>
From the class website of Professor Xe. I found it very clear.
 - Garcia B. 2013. Implementation of a double-hurdle model. *Stata Journal* **13**(4): 776-794

I doubt we'll discuss the double-hurdle model, given time constraints, but this article does a good job of introducing it generally and per a new command in Stata.
- Next time
 - Some notes on tobit
<https://www.stata.com/statalist/archive/2012-02/msg00598.html>
 - Holden D (2004) "Testing the Normality Assumption in the Tobit Model." *Journal of Applied Statistics*. 31(5):521-532.
 - Barros, M., Galea, M., González, M., & Leiva, V. (2010). Influence diagnostics in the tobit censored response model. *Statistical Methods & Applications*, 19(3), 379-397.
http://staff.deuv.cl/leiva/archivos/leiva_art/barros_galea_gonzalez_leiva_2010.pdf

Holden, D. (2011). Testing for heteroskedasticity in the tobit and probit models. *Journal of Applied Statistics*, 38(4), 735-744.

Vincent, D. (July 22, 2010). BTOBIT: Stata module to produce a test of the tobit specification. <http://ideas.repec.org/c/boc/bocode/s457163.html>

Drukker, D.M. (2002). Bootstrapping a conditional moments test for normality after tobit estimation. *The Stata Journal*, 2(2), 125-139. <http://www.stata-journal.com/sjpdf.html?articlenum=st0011>

Newey, W.K. (1987). Specification tests for distributional assumptions in the Tobit model. *Journal of Econometrics*, 34(1-2), 125-145.

Wilhelm, M.O. (2008). Practical Considerations for Choosing Between Tobit and SCLS or CLAD Estimators for Censored Regression Models with an Application to Charitable Giving. *Oxford Bulletin of Economics and Statistics*, 70(4), 559-582.

I also find the following three papers useful as references.

S. B. Caudill and F. G. Mixon. **More on testing the normality assumption in the Tobit Model.** *Journal of Applied Statistics*, 36(12):1345-1352, 2009. D. Holden. **Testing the Normality Assumption in the Tobit Model.** *Journal of Applied Statistics*, 31(5):521-532, 2004. J. B. McDonald and H. Nguyen. **Heteroscedasticity and Distributional Assumptions in the Censored Regression Model.** *Communications in Statistics - Simulation and Computation*, 44(8):2151-2168, 2014.

Assignment

- Assignment 13: Count data models

Session 14 Sample selection

Apr 17, 2017

Now we move to the *other* type of sample selection, where data is missing as a function of something other than the value of the dependent variable. As you'll read, at least two things are frequently conflated here. I've decided to discuss both today, partially so we can be sure about the distinction between them.

Readings

I would start with the Certo *et al* paper, which does a very clear job of explaining distinctions between the various issues often conflated under "sample selection." Then read Shaver, which is a foundation citation in this area. Hamilton & Nickerson provide a nice integration across the first two papers. The Heckman model featured in these pages has been criticized on various grounds. Puhani is a heavily cited paper in that more critical tradition. Lastly, turn to the two papers on the propensity score method, which address a slightly different problem.

- Method

- Certo ST, Busenbark JR, Woo H-S, Semadeni M. 2015. Sample selection bias and heckman models in strategic management research. *Strategic Management Journal*: n/a-n/a
- Hamilton BH, Nickerson JA. 2003. Correcting for endogeneity in strategic management research. *Strategic Organization* **1**(1): 51-78
- Li M. 2012. Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods* **16**(2): 188-226
- Puhani PA. 2000. The heckman correction for sample selection and its critique. *Journal of Economic Surveys* **14**(1): 53-69
- Shaver JM. 1998. Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect fdi survival? *Management Science* **44**(4): 571-585

- Use

- Berry H, Kaul A. 2015. Global sourcing and foreign knowledge seeking. *Management Science* **61**(5): 1052-1071
- Castañer X, Mulotte L, Garrette B, Dussauge P. 2014. Governance mode vs. Governance fit: Performance implications of make-or-ally choices for product innovation in the worldwide aircraft industry, 1942-2000. *Strategic Management Journal* **35**(9): 1386-1397

- Dushnitsky G, Shapira Z. 2010. Entrepreneurial finance meets organizational reality: Comparing investment practices and performance of corporate and independent venture capitalists. *Strategic Management Journal* **31**(9): 990-1017
- Gomulya D, Boeker W. 2014. How firms respond to financial restatement: Ceo successors and external reactions. *Academy of Management Journal* **57**(6): 1759-1785
- Sampson R. 2004. The cost of misaligned governance in R&D alliance. *Journal of Law, Economics and Organization* **20**(2): 484-526
- Additional references
 - Heckman JJ, Lopes HF, Piatek R. 2014. Treatment effects: A Bayesian perspective. *Econometric Reviews* **33**(1-4): 36-67
 - Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* **47**(1): 153-161
 - Heckman JJ. 2005. The scientific model of causality. *Sociological Methodology* **35**(1): 1-97
 - Imbens GW, Wooldridge JM. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* **47**(1): 5-86
 - Practically encyclopedic reference to everything we've talked about in this module and more.
 - Imbens GW. 2015. Matching methods in practice: Three examples. *Journal of Human Resources* **50**(2): 373-419
 - Up to date overview of propensity score and other matching techniques.
 - Lennox CS, Francis JR, Wang Z. 2012. Selection models in accounting research. *Accounting Review* **87**(2): 589-616
 - Stolzenberg RM, Relles DA. 1997. Tools for intuition about sample selection bias and its correction. *American Sociological Review*: 494-507
 - Vella F. 1998. Estimating models with sample selection bias: A survey. *Journal of Human Resources*: 127-169

Assignment

- Assignment 14 - Truncation and censoring

Session 15 Endogeneity

Apr 25, 2017

So far in this module, our problems have been discrete. Observations either are in the sample or missing. Firms choose a specific strategy. Today, we'll move into the world of continuous problems, otherwise known as instrumental variables. For example, in examining the relationship between R&D spending and profitability, we may wish to account for the possibility that there are unobserved variables that affect both R&D spending and profitability of a firm. If these exist, our results may be biased. The difference between today's topic and what's come before is not actually that definitive. For example, some scholars argue that instrumental variables are more appropriate than the models discussed in the last class, even when the non-randomly determined independent variable is discrete. An additional opportunity comes about because the methods most often used in these situations, two- and three-stage least squares, are also used in the related, but distinct, circumstance of simultaneity. We won't spend much if any time on simultaneous equations in this class, but you'll see them discussed as part of your readings. Just for your elucidation, the classic example of simultaneity is supply and demand. Economic models assume that in a perfectly competitive market, the price of a product will naturally settle on the price that will clear the market—that is, make demand and supply balance. But, price determines both how much consumers demand and how much firms supply. Thus, supply and demand are simultaneously determined, which makes them impossible to estimate with conventional methods.

Reading

Start with Kennedy and Wooldridge to get a background. Basile bridges the general concept to its application in strategy. Semadeni demonstrates through simulation the impact endogeneity can have and the varied effectiveness of proposed solutions. Then skim Antonakis. I include it for two reasons, besides its general clarity. First, it has become a touchstone you should be aware of. Second, it does a good job of comparing approaches to various flavors of non-exogeneity, not just what we've discussed in this module, but more broadly.

I've included several papers that sound a cautionary note about how endogeneity is often modeled. We'll split up the reading of these. Giles is a short, but important piece, focusing on the seemingly technical issue of whether the standard error of 2SLS estimators is identified or not. Deaton takes a rather skeptical view of the whole state of affairs. His article actually more directly focused on the type of sample selection problems we discussed last week, but I think his philosophical concerns are generally applicable to all types of instrumental variable estimation. Dunning explores an often under-appreciated assumption, that variation in the endogenous regressor that is related to the instrumental variable has the same effect as variation that is unrelated to the instrument. Qin "exposes the flaw in defining endogeneity bias by correlation between an explanatory variable and the error term of a regression model."

Lastly, skim Chatterji *et al* for a totally different approach to concerns about endogeneity, conducting field experiments.

- Method

- Antonakis J, Bendahan S, Jacquart P, Lalive R. 2010. On making causal claims: A review and recommendations. *Leadership Quarterly* **21**(6): 1086-1120
- Bascle G. 2008. Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization* **6**(3): 285-327
- Chatterji AK, Findley M, Jensen NM, Meier S, Nielson D. 2016. Field experiments in strategy research. *Strategic Management Journal* **37**(1): 116-132
- Deaton A. 2010. Instruments, randomization, and learning about development. *Journal of Economic Literature* **48**(2): 424-455
- Dunning T. 2008. Model specification in instrumental-variables regression. *Political Analysis* **16**(3): 290-302
- Giles, Dave. 2014 Finite-sample properties of the 2SLS estimator. *Econometrics Beat: Dave Giles' Blog* July 11.
- Kennedy, Chapter 9, "Violating Assumption Four: Instrumental Variable Estimation"
- Qin D. 2015. Time to demystify endogeneity bias.
- Semadeni M, Withers MC, Trevis CS. 2014. The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal* **35** (7): 1070-1079
- Wooldridge, Chapter 15 Instrumental Variables Estimation and Two Stage Least Squares

- Use

- Hoetker G, Mellewigt T. 2009. Choice and performance of governance mechanisms: Matching alliance governance to asset type. *Strategic Management Journal* **30**(10): 1025-1044
- Sanders WG, Hambrick DC. 2007. Swinging for the fences: The effects of ceo stock options on company risk taking and performance. *Academy of Management Journal* **50**(5): 1055-1078

- Additional references

- Angrist JD, Pischke J-S. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press

This book has become a "go-to" source on quasi-experimental approaches to econometrics.

- Bollen KA. 2012. Instrumental variables in sociology and the social sciences. *Annual Review of Sociology* **38**(1): 37-72
- Crespo-Tenorio A, Montgomery JM. 2013. A Bayesian approach to inference with instrumental variables: Improving estimation of treatment effects with weak instruments and small samples.
- Hamilton BH, Nickerson JA. 2003. Correcting for endogeneity in strategic management research. *Strategic Organization* **1**(1): 51-78
- Heckman JJ, Urzua S. 2010. Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics* **156**(1): 27-37
- Imai K, Keele L, Tingley D, Yamamoto T. 2011. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review* **105**(04): 765-789
- Larcker DF, Rusticus TO. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* **49**(3): 186-205
- Lee DS, Lemieux T. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* **48**(2): 281-355

We may or may not have time to discuss regression discontinuity models, but you should be aware of them at least. RD is another attempt to control for non-random treatment. To vastly simplify, imagine you want to study the effect of attending a prestigious college on subsequent income. There is probably little *a priori* difference between the lowest ranked (on some relevant criteria) person who got into Harvard *etc.* and the highest ranked person who didn't get in. So, subsequent differences in the person who just scraped in and the person who just missed should be due to the experience of attending Harvard. That's the basic idea. This paper gives a good introduction to the recent state of the literature.

- Lopes HF, Polson NG. 2014. Bayesian instrumental variables: Priors and likelihoods. *Econometric Reviews* **33**(1-4): 100-121
- Petrin A, Train K. A control function approach to endogeneity in consumer choice models. *Journal of marketing research* **47**(1): 3-13
- Wooldridge JM. 2015. Control function methods in applied econometrics. *Journal of Human Resources* **50**(2): 420-445

Control functions as a response to endogeneity haven't gotten a lot of use in strategy and I see limitations in their use. Nevertheless, they are an important technique and this is an accessible and current introduction.

Assignment

- Assignment 15: Sample selection

Conclusion and finals

Pseudo-session 16 Conclusion and finals

As detailed in the syllabus, this class has a final paper and final exam. There is also a weekly assignment reviewing the material from session 15.

Assignment

- Assignment 16: Endogeneity