SOCI/DEMG/CRIM 662: Panel Data & Causal Inference

Professor Xi Song

Fall, 2022

[version 6/9/2022]

E-mail: xisong@upenn.edu
Office Hours: TBA
Office: McNeil 271

Web: **https://canvas.upenn.edu/courses/1533157**
Class Hours: Monday/Wednesday 1:30–2:45pm
https://xisong-soc662fall2022.youcanbook.me/

Allison Dunatchik
Office location: TBA
Lab location: TBA

TA Email: adunat@sas.upenn.edu
Office hours: TBA (for weeks with no labs)
Lab Hours: TBA

Course Overview

Panel data or longitudinal data consist of multiple measures over time on a sample of individuals. These types of data occur extensively in both observational and experimental studies in social, behavioral, and health sciences. This course will provide an introduction to the principles and methods for the analysis of panel data. Whereas some supporting statistical theory will be given, emphasis will be on data analysis and interpretation of models for longitudinal data. Problems will be motivated by applications primarily in social sciences.

Prerequisites

A prior statistics course—SOCI 536, or the equivalent—is required. Prior experiences with data analyses using *R* are highly recommended but not required. Previous exposure to (or use of) causal inference methods (e.g., counterfactual thinking and the potential outcomes framework) is required because we will focus on causal inference in a longitudinal setting.

Contacts

You can reach me via email; however, I do not respond to email between 9 pm and 9 am (and neither do the teaching assistant) or over the weekend. If I don't respond within 24 hours, please feel free to send me a polite reminder. I don't intend to be unavailable, but sometimes I get quite a lot of email and/or I simply get swamped. Reminders do not offend me.

I will respond to most of the emails regarding the course, and this is the best way to work through simple questions. Please check your email and Canvas/Piazza several times a week. Email is one of the best ways to keep in touch with our class when we are not in class. More complex questions would likely require more time, and for these, I recommend my office hours.

Textbooks

- **Required**

1. Singer, Judith D., John B. Willett, and John B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. NY: Oxford University Press. (Hereafter Singer & Willett)
2. Hernan, Miguel and James Robins. 2020. *Causal Inference: What If.* Book [Open access]

- **Optional**

3. Mills, Melinda. 2011. *Introducing Survival and Event History Analysis*. London: Sage Publication. (illustration using R)
4. Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods* (Second Edition). CA: Sage Publications.
5. Nagin, Daniel S. 2005. *Group-Based Modeling of Development.* Boston, MA: Harvard University Press.

- **Recommended**

6. Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.
7. Hedeker, Donald, and Robert D. Gibbons. 2006. *Longitudinal Data Analysis*. NY: John Wiley & Sons.

Required Software

- **R**
- **RStudio** is an integrated development environment for **R**, a programming language for statistical computing and graphics.
  - **R** tutorial courses on **DataCamp**
  - More tutorials in TA sessions
- LaTex document processors, such as **Overleaf**, **LyX**, **TeXstudio**, or **Knitr**.

Class Requirements and Evaluation

1) Bi-Weekly Problem Sets (30% of your final grade)

Problem sets will be due in class the week after they are assigned on **Wednesday**. Please submit an electronic version to Canvas. Any programming language is accepted for the simulation exercises. **If students have any questions on Problem Sets they should first ask TA and only ask the professor if the TA is unable to help.**

There are **4** assignments in the semester (except for the two midterm weeks). It is important that you do each set of weekly assignment completely and on time; **late submissions will not be accepted**. If for some reason you do not complete your assignment on time, I encourage you to complete it on your own, but we will not accept it for credit. In the first few weeks, you will be doing analyses using a major U.S. national sample survey (e.g., NLSY79). As the semester progresses, however, for most of the assignments you will be able to substitute data of your own, focusing on topics that interest you and/or that pertain to your term paper.

2) Two Midterm Exams (30% of your final grade)

The mid-term exams will take place on **October 5, 2022** and **November 16, 2022** from 1:30 pm to 2:45 pm. The exams are open-book.

3) Final Paper (30% of your final grade) & Presentation at the Mini-Conference (10% of your final grade)

The course will culminate in a term paper on a topic of your choosing in which you will carry out a quantitative analysis of some substantive issue using the technical and analytic skills developed by doing the assignments. It is not uncommon for course term papers to lead to or revise master's papers or chapters of Ph.D. dissertations and/or publications.

Each student is required to present their final paper in the mini-conference during the last week of the class.

With instructor's prior approval, you may write co-authored papers with **no more than two** authors. Both authors must be students in the class. In the case of co-authorship, the paper should detail what each author contributes to the project and include a separate paragraph or document detailing what each author contributed.

Your final term paper will be due at the end of the semester on **December 22, 2022 (Thursday), 5 pm**. Late papers will not be accepted. More information on this project will be distributed over the semester.

Course Policies

During Class

I understand that the electronic recording of notes will be important for class and so computers will be allowed in class. I will upload my notes to Canvas after each class. Please refrain from using computers for anything but activities related to the class. Phones are prohibited as they

are rarely useful for anything in the course. Eating and drinking are allowed in class but please refrain from it affecting the course.

After Class

I expect you to spend **3** hours in class and **4-8** hours after class each week on the course subjects.

Office Hours

The scheduled office hours are on TBA. If, for some weeks, I have to cancel my office hours, I will email the class about my new office hours or you can send me an email to schedule separate meetings.

Policies on Incomplete Grades and Late Assignments

Late assignments will not be accepted. See the policy discussed about weekly problem sets.

Grading Policies

The typical UPenn grading scale will be used. Normally, grading will not be on a curve. You can access your personal grades on the course web page as we move along in the course. Your final course grade will be figured according to the following cutoffs:

A = 94 – 100	C = 73 – 76
A- = 90 – 93	C- = 70 – 72
B+ = 87 – 89	D+ = 67 – 69
B = 83 – 86	D = 63 – 66
B- = 80 – 82	D- = 60 – 62
C+ = 77 – 79	F = 59 and Below

However, if no one receives higher than 90+, I reserve the right to curve the scale dependent on overall class scores at the end of the semester. Any curve will only ever make it easier to obtain a certain letter grade.

Canvas

You can download all course materials from the course Canvas website:

https://canvas.upenn.edu/courses/1533157

Plan of Lecture

- Basic Concepts & Introduction
- An Introduction to Growth Curve Models, I
- Growth Curve Models, II
- Covariance Pattern Models & GEE
- Bayesian Growth Curve Models
- First Midterm Examination (Open Book)
- Introduction to Causal Inference & Causal Diagrams
- IPTW & Marginal Structural Models
- Structural Nested Mean Models
- Machine Learning in Causal Inference
- Second Midterm Examination (Open Book)
- Group-Based Trajectory Models & Dual Trajectory Analysis
- Sequence Analysis & Dyadic/Polyadic Sequence Analysis
- Mini-Conference

Class Schedules (Subject to Change)

Topic 1 (August 31): *Basic Concepts & Introduction*

- Chapter 1 in Singer & Willett.
- Chapter 2 in Singer & Willett.

September 4 No Class (Labor Day)

Topic 2 (September 7): *An Introduction to Growth Curve Models*

- Chapters 3, 4 in Singer & Willett.
- Chapters 1, 2 in Raudenbush & Bryk,

Topic 3 (September 12 & 14): *Growth Curve Models II*

- Exercise 1 due
- Chapter 5 in Singer & Willett.
- Chapters 3, 4, 5 in Raudenbush & Bryk
- Jæger, Mads Meier. 2012. "The Extended Family and Children's Educational Success." *American Sociological Review* 77, no. 6: 903-922.
- Huffman, Matt L., and Philip N. Cohen. 2004. "Racial Wage Inequality: Job Segregation and Devaluation across US Labor Markets." *American Journal of Sociology* 109, no. 4: 902-936.

Topic 4 (September 19 & 21): *Covariance Pattern Models & GEE*

- Chapter 7 in Singer & Willett.
- Chapter 6 in Gibbons & Hedeker
- Ballinger G.A. (2004). Using generalized estimating equations for longitudinal data analysis, *Organizational Research Methods*, 7:127-150.
- Diggle P.J., Heagerty P., Liang K.-Y., Zeger S.L. (2002). *Analysis of Longitudinal Data*, 2nd edition, New York: Oxford University Press.
- Dunlop D.D. (1994). Regression for longitudinal data: a bridge from least squares regression, *The American Statistician*, 48:299-303.
- Hardin J.W., Hilbe J.M. (2003). *Generalized Estimating Equations*, New York: Chapman and Hall.
- Hu F.B., Goldberg J., Hedeker D., Flay B.R., Pentz M.A. (1998). A comparison of generalized estimating equation and random-effects approaches to analyzing binary outcomes from longitudinal studies: illustrations from a smoking prevention study, *American Journal of Epidemiology*, 147:694-703.
- Norton E.C., Bieler G.S., Ennett S.T., Zarkin G.A. (1996). Analysis of prevention program effectiveness with clustered data using generalized estimating equations, *Journal of Consulting and Clinical Psychology*, 64:919-926.
- Zorn C.J.W. (2001). Generalized estimating equation models for correlated data: a review with applications, *American Journal of Political Science*, 45:470-490.

Topic 5 (September 26 & 28): *A Bayesian Approach to Growth Curve Models*

- Exercise 2 due

First Midterm (October 3 & 5): ***Review & Examination (Open Book)***

October 6–9 No Class (Fall Break)

Topic 6 (October 10 & 12): ***Introduction to Causal Inference & Causal Diagrams***

- Chapters 1 & 3 & 6 in Hernan & Robin (2020)

Topic 7 (October 17 & 19): ***Confounding, Selection Bias, and Measurement Bias***

- Chapters 7 & 8 & 9 in Hernan & Robin (2020)

Topic 8 (October 24 & 26): ***IPTW & Marginal Structural Models***

- Exercise 3 due
- Chapters 11 & 12 in Hernan & Robin (2020)

Topic 9 (October 31 & November 2): ***Standardization and the Parametric G-formula***

- Chapter 13 in Hernan & Robin (2020)

Topic 10 (November 7 & 9): ***G-estimation of Structural Nested Mean Models***

- Exercise 4 due
- Chapter 14 in Hernan & Robin (2020)

Second Midterm (November 14 & 16): ***Review & Examination***

Advanced Topic (November 21 & 28): ***Group-Based Trajectory Models & Dual Trajectory Analysis***

- Guest Speaker: Allison Dunatchik
- Nagin (2005)
- Jones, Bobby L. and Daniel S. Nagin. 2007. "Advances in Group-Based Trajectory Modeling and a SAS Procedure for Estimating Them." *Sociological Methods & Research* 35:542–571.
- Warren, John Robert, Liying Luo, Andrew Halpern-Manners, James M. Raymo, and Alberto Palloni. 2015. "Do Different Methods for Modeling Age-Graded Trajectories Yield Consistent and Valid Results?." *American Journal of Sociology* 120, no. 6: 1809-1856.
- Qian, Yue. 2018. "Educational Assortative Mating and Income Dynamics in Couples: A Longitudinal Dyadic Perspective." *Journal of Marriage and Family* 80(3): 607–621.

- Keizer, Renske, and Niels Schenk. 2012. "Becoming a Parent and Relationship Satisfaction: A Longitudinal Dyadic Perspective." *Journal of Marriage and Family* 74(4): 759–773.

September 23 No Class (Thanksgiving)

Advanced Topic (November 30 & December 5): *Sequence Analysis & Dyadic/Polyadic Sequence Analysis*

- Chapter 11 in Mills (2011)
- Aisenbrey, Silke, and Anette E. Fasang. 2010. "New Life for Old Ideas: The 'Second Wave' of Sequence Analysis Bringing the Course Back into the Life Course." *Sociological Methods & Research* 38.3: 420–462.
- Liao, Tim F. 2021. "Using Sequence Analysis to Quantify How Strongly Life Courses Are Linked." *Sociological Science* 8: 48–72.

Advanced Topic (December 7 & 12): *Mini-Conference*

- More information about the mini-conference will be posted later.

Advanced Topics

We will cover advanced topics on group-based trajectory models, sequence analysis, and variants of these models when two or more outcomes are jointly modelled (known as dyadic/polyadic or bi-/multi-variate models). We will invite guest speakers to show applications of these methods in their own research.

Student Presentation Conference

Overview

All students or research teams will need to give a final presentation about their research projects during the last class on **April 26 3:30 om to 6:30 pm**. We will divide all class presentations into 3-4 sessions. Each session will include 1 session chair, 1-2 discussants, and 4 presenters. Presenters will be asked to send their final project proposal to the session chair, discussant(s) and other presenters well in advance of the session.

Presentation Preparation

It is imperative that discussants receive your proposal no later than **April 22, 2022**, so there is time to read all papers and prepare comments; some chairs/discussants may allow a later date, but this must be directly arranged and not assumed. Please be sure to send your proposal to the session chair and other presenters when you send it to the discussant(s); this will enhance the exchange of ideas within the session.

Time

Following the rule of PAA, most presentations will only be 12 minutes long. The chair of your session will inform you of the exact time allocated to your presentation; if you have not heard from the session chair, assume that you will have 12 minutes. The time you are allotted will not include Q&A unless otherwise indicated by your chair; typically, session chairs reserve 10 or so minutes at the end of each session for questions and discussion.

Plan For Your Presentation

Below are some presentation tips provided by PAA. A good conference presentation provides a clear and succinct overview of your paper. Consider the time available and the multiple learning styles of attendees (auditory, visual, etc.) to create a valuable presentation.

- *Prepare visual aids*: Most presenters use slides, either in PowerPoint or PDF, as visual aids for their presentation.
- *Type*: Use at least 20-point type so that audience members can easily read the print on your slides. Please do not include large tables in your slides: Summarize your key results rather than presenting large, dense tables.
- *Bullets*: Limit yourself to 3-4 bullets per slide and 10 or fewer words per bullet. *Number*: A rough rule of thumb is to prepare no more than one slide for every minute you will be presenting.
- Try to avoid the use of acronyms, jargon, and abbreviations: Past conference evaluations have clearly indicated that one frustration, in particular for new and international attendees, is the use of 'insider' language, acronyms, and abbreviations that make it difficult to comprehend a presentation.

- Consider livening up your slides with graphics and pictures: Graphics can be very effective in capturing the audience's attention and focusing them on the point you want to make.
- Contact information slide: Include a slide that you put up at the beginning with your presentation title, name, and contact information.
- Please proof read and spell check.
- Practice: Practice your presentation to ensure that it highlights key points, your delivery is clear, and you finish within the time allocated.
- Email your slides: Email your slides to the session chair in advance of the session in case there are any difficulties with screen sharing.

Final Paper Instruction

You will develop a research paper during the semester. Instead of just memorizing methodological concepts and techniques, you should be able to apply them to an actual research project. The paper should include the objective of the research, a short literature review and one or two hypotheses, variables for study, and measurement. During the latter part of the quarter, you will have to conduct data analyses to test your hypothesis or support your argument. You can either bring your own social science data or use one module of General Social Surveys for this research paper.

I strongly encourage you to discuss your research ideas and plan of analysis with me as early as possible during the semester. You are also encouraged to work with a classmate (no more than 2 authors on a paper) on this final project. In fact, a lot of my own work is collaborative and several are derived from my collaborations with classmates during graduate school (the papers were not published until many years later or still unpublished). In the case of co-authorship, the paper should detail what each author contributes to the project and include a separate paragraph or document detailing what each author contributed.

The final paper should be typed and at most **10** pages of 12 pt, including any text, tables, and figures. **I will only read the first 10 pages if your paper exceeds the page limit.** You should also include references, R/Stata commands, and appendices, but these materials are not counted toward the page limit. Please include margins of at least one inch on all sides of the paper. **It should be submitted via the Canvas website and delivered to my mailbox in McNeil 353**.

The report should be self-contained and suitable for a non-statistician with a college level of knowledge of statistics.

Advanced modeling methods should be defined briefly in the text to the extent necessary for understanding of the results, along with a reference. By advanced I mean methods that were not prerequisite for the course.

It is recommended that the following outline be followed in preparing the report:

1. *Abstract*. This should consist of a brief statement of the results of your analysis. This should be like that of a research paper analyzing the data.

2. *Introduction*. Here includes a clear statement of the scientific questions addressed by your analysis of the data. The goal of the statistical analysis and the social context of the problem should be clear to all who read the introduction.

3. *Analysis and Results*. Describe your analysis and its results clearly and concisely. If necessary, use graphical displays and tables to convey the results. State clearly what your research and null hypotheses and underlying assumptions (in terms of variable type, distribution, sampling, target population, etc) are. Describe methods used, approaches taken to examine the underlying assumptions and so on. Explain why the methodology is appropriate.

4. *Discussion*. This section should describe the sociological and statistical issues raised by the results described in the previous section. Limitations of the study and of the analysis should be discussed here. There may be social policy issues that you would have liked to discuss with the investigator if this had been a real collaboration with a non-statistical social scientist. If so, describe these issues and why they would be relevant to the analysis and/or interpretation of results. If appropriate, provide suggestions for further analysis or collection of additional data. Summarize your conclusions about the issue of sociological concern.

5. *References*. List the key books and articles you consulted that are reflected in your report, as well as references that might be useful to the reader if they want to know more.

6. *Tables and Figures*. These should be included in the text, but can also be included in a separate section at the end. All should have clear titles/captions, and figures should have explanatory legends.

7. *Appendices*. There can be one or more appendices. You may include more technical discussion of your methodology or any theory developed to implement your models. Appendices are not required and should be included only if you feel they add something important to your report.

You should spend the majority of your time thinking about the sociological and statistical issues and writing the report rather than spending all your time carrying out the statistical analysis. You may do a wonderful job of analysis, but it is of no use unless you can communicate the results to your audience.

If you decide to write your paper in LaTeX, you can use my template.

Evaluation

Criteria used to judge performance will include the following three factors, each given equal weight:

1. *Statistical appropriateness*. Appropriateness of the analysis and models for the data and questions. Technical execution of the analysis.

2. *Scientific appropriateness*. Thoughtfulness and simplicity of your analysis. Does your analysis really answer the sociological questions of interest?

3. *Quality of the written report*. The report will be judged based upon its organization, clarity, and accuracy. Simple, concise sentences are preferred to sentences that are convoluted or otherwise confusing.

You can find information about writing and editing service provided by the university in the course syllabus.

Choice of your data set

If you decide to use your own data, the data set should contain at least 30 cases, and at least four variables measured on each case. At least one of the variables should be an outcome variable that you want to predict or model via the other variables. All data sources should be cited and described. If you have questions about the selection of data, you can come to talk with me. Do you merely use examples from the lectures—the social world is an interesting place!

Your statistical task is to model the structure in the data and describe it. To do this, you should consider the various forms of social structure considered in the course.

Here are some suggestions:

- Start by presenting visual summaries of the data, followed by numerical summary measures.
- Describe the impact of the covariates on the outcome variable(s) via simple models first.
- For any model be sure to consider diagnostics of the appropriateness of the model.
- Construct summaries of the overall quality of the fit, and a comparison between alternatives to the model finally used.

This is an open-ended question, so feel free to experiment.

Additional Recommended Readings by Topic

The following list of readings offers further elaboration or additional topics on Panel Data Analysis beyond the course. The list provides you with references to pursue as you advance the research design for your term paper or to elaborate on a week's topic. (Note: There may be newer editions to some of these books.)

Causal Inference

- Bastias, Pablo Geraldo, and Jennie E. Brand. 2020. Causal Inference. Oxford University Press.

  [for an overview of the causal inference literature and its relevance to sociology;] https://www.pablogeraldo.com/research/publications/GeraldoBrand_2020_Causality.pdf

- Elwert, Felix, and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40: 31–53.

- Gangl, Markus. 2010. "Causal Inference in Sociological Research." *Annual Review of Sociology* 36:21–47.

- Hong, Guanglei. 2015. *Causality in a Social World: Moderation, Mediation and Spill-over*. John Wiley & Sons.

- Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

- Manski, Charles. 1995. *Identification Problems in the Social Sciences*. Boston, MA: Harvard University Press.

- Morgan, Stephen, ed. 2013. *Handbook of Causal Analysis for Social Research*. Springer Series.

- Morgan, Stephen, and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research* Second Edition. New York, NY: Cambridge University Press.

- Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. New York, NY: Cambridge University Press.

- Rosenbaum, Paul. 2010. *Observational Studies*. New York: Springer.

- Rubin, Donald. 2006. *Matched Sampling for Causal Effects*. New York, NY: Cambridge University Press.

- VanderWeele, Tyler. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

- Huber, Martin. 2021. *Causal Analysis: Impact Evaluation and Causal Machine Learning with Applications in R* [Open Access]

- Huntington-Klein, Nick. 2021. *The Effect*. CRC Press.

- Cunningham, Scott. 2020. *Causal Inference: The Mix Tape.* [Open Access]
- Xu, Yiqing. Causal Inference with Panel Data. Youtube **videos** [A review of the synthetic control method]

Sample Selection Bias

- Berk, Richard A. 1983. "An Introduction to Sample Selection Bias in Sociological Data." *American Sociological Review* 48:386–98.
- Breen, Richard. 1996. *Regression Models: Censored, Sample Selected, or Truncated Data*. (Sage No. 111)
- Kang Fu, Vincent, Christopher Winship, and Robert D. Mare. 2004. "Sample Selection Bias Models." Pp. 409–430 in *Handbook of Data Analysis*, Melissa Hardy and Alan Bryman, eds., Sage Publications. [An update to the citation below by Winship and Mare]
- Winship, Christopher, and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology* 18:327–50.

Basic Econometrics

- Angrist, Joshua and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Greene, William H. 2008. *Econometric Analysis* 6th Edition. Upper Saddle River, NJ: Prentice Hall.
- Woolridge, Jeffrey M. 2002. *Econometric Analysis of Cross-Section and Panel Data*. Cambridge: The MIT Press.
- Wooldridge, Jeffrey M. 2019. *Introductory Econometrics: A Modern Approach*. 7th edition. Cengage Learning.

Fixed and Random Effects Models

- Allison, Paul. 1990. "Change Scores as Dependent Variables in Regression Analysis." *Sociological Methodology* 20:93–114.
- Allison, Paul. 1994. "Using Panel Data to Estimate the Effects of Events." *Sociological Methods and Research* 23:174–199. Baltagi, Badi H. 2008. *Econometric Analysis of Panel Data* 4th Edition. New York: Wiley.
- Halaby, Charles N. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30:507–44. [A nice review, focusing on causal inference and fixed effects models;]
- Hausman, Jerry A. 1978. "Specification Tests in Econometrics." *Econometrica* 46:1251–1272. [Hausman tests are frequently used to compare fixed and random effects models; this article introduces the test;]

- McNeish, Daniel, and Ken Kelley. 2019. "Fixed Effects Models Versus Mixed Effects Models for Clustered Data: Reviewing the Approaches, Disentangling the Differences, and Making Recommendations." *Psychological Methods* 24(1): 20. [Compares differences between Econometric and Sociological/Psychological approaches to handling clustered/longitudinal data]

- Lambert, Ben. 2013. "Panel Data Econometrics—An Introduction." YouTube Video.11:01. October 4, 2013. [See the succeeding **videos** in the sequence for a full introduction to fixed effects models]

Propensity Score Matching

- Abadie, Alberto, David Drukker, Jane Leber Herr, and Guido Imbens. 2002. "Implementing Matching Estimators for Average Treatment Effects in Stata." *Stata Journal*. [A nice Stata routine to use for PS matching.]

- Becker, Sascha O., and Andrea Ichino. 2002. "Estimation of Average Treatment Effects Based on Propensity Scores." *Stata Journal* 2:358–377. [Use this routine to generate p-scores. It also contains matching estimators;]

- Morgan, Stephen and David Harding. 2006. "Matching Estimators of Causal Effects." *Sociological Methods & Research* 35(1):3–60.

- Rosenbaum, Paul R. and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies using Sub-classification on the Propensity Score." *Journal of the American Statistical Association* 79, 516–524.

- Smith, Herbert L. 1997. "Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies." *Sociological Methodology* 27:325–353.

Instrumental Variables

- Angrist, Joshua. D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434): 444–455.

- Angrist, Joshua and Alan Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15(4): 69–85.

- Heckman, James. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32(3): 441–462.

- Imbens, Guido and Joshua Angrist. 1997. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62: 467–476.

Growth Curve and Trajectory Models

- Bollen, Kenneth A. and Patrick J. Curran. 2006. *Latent Curve Models: A Structural Equation Perspective.* New York: Wiley.

Longitudinal Network Analysis

- Krivitsky, Pavel N., and Mark S. Handcock. 2014. "A Separable Model for Dynamic Networks." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(1): 29-46.