

SOCI/DEMG/CRIM 662: Panel Data Analysis

Professor Xi Song

Fall, 2020

[version 9/9/20]

E-mail: xisong@upenn.edu

Office Hours: Monday 10am–12pm

Office: McNeil 271

Office Hour Zoom ID: [972 0678 2838](#)

Web: <https://canvas.upenn.edu/courses/1533157>

Class Hours: Thursday 3–5:50pm

Zoom Class ID: [938 3457 9505](#)

<https://xisong-soc662fall2020.youcanbook.me/>

Nick Graetz

TA Email: ngraetz@sas.upenn.edu

Office hours Zoom ID: [983 3776 4924](#)

Lab Zoom ID: [996 0791 5116](#)

Office Hours: Friday 1pm–2pm

(12pm–2pm for weeks with no labs)

Lab Hours: Friday 12pm–1pm

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.

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)

- **Optional**

2. Allison, Paul D. 2010. *Survival Analysis Using SAS: A Practical Guide*. SAS Institute.
3. Allison, Paul D. 2014. *Event History and Survival Analysis: Regression for Longitudinal Event Data*. Vol. 46. Sage publications.
4. Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press, 2006.
5. Hedeker, Donald, and Robert D. Gibbons. 2006. *Longitudinal Data Analysis*. NY: John Wiley & Sons.
6. Mills, Melinda. 2011. *Introducing Survival and Event History Analysis*. London: Sage Publication. (illustration using R)
7. Cleves, Mario, William Gould, William W. Gould, Roberto Gutierrez, and Yulia Marchenko. 2016. *An Introduction to Survival Analysis using Stata*. Stata press.
8. Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods* (Second Edition). CA: Sage Publications.
9. Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Boston, MA: MIT press.
10. Wooldridge, Jeffrey M. 2019. *Introductory Econometrics: A Modern Approach*. 7th edition. Cengage Learning. (Chapters 13, 14)

- **Recommended**

11. Long, J. Scott. 2009. *The Workflow of Data Analysis using Stata*. Stata Press.
 Note: Although this book focuses on Stata rather than R, the author offers many useful suggestions about file organizations, planning, and replications.
12. Miller, Jane E. 2005. *The Chicago Guide to Writing about Multivariate Analysis*. University of Chicago Press.

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 (40% of your final grade)

Problem sets will be due in class the week after they are assigned on **Thursday** at 3 pm. 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 5 assignments in the semester (except for the midterm week and the final week). They will be read and returned the following week on **Thursdays**. 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. **To compensate for this strict policy, I will drop the lowest grade you receive on an assignment when we tabulate your overall grade.** In the first few weeks, you will be doing analyses using a major U.S. national sample survey (e.g., NLSY79). As the quarter 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) Midterm Exam (20% of your final grade)

The mid-term exam will take place on **October 15, 2020** (or the next Thursday after we finish **Covariance Pattern Models**) from 3 pm to 6 pm. The exam is open-book.

3) Final Paper (and Term Paper Proposal) (40% 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.

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 20, 2020 (Sunday), 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 at least **8** hours after class each week on the course subjects.

Office Hours

The scheduled office hours are on Mondays from 11 am to 1 pm. If, for some weeks, I have to cancel my office hours on Mondays, 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.

A Guide for Emergency & Stress

- If you have an emergency, call 511 or the Division of Public Safety 215-573-3333. Help is available 24/7.
- If you need consultation, call CAPS (Counseling and Psychological Services): 215-898-7021. On call clinicians are available 24/7.
- If you have academic difficulty, please contact me or seek help from the following services:
 1. **Weingarten Learning Resources Center**: Professional instruction in skills such as academic reading, test taking, and study strategies.
 2. **Tutoring Center**: A variety of subject-specific peer tutoring services to supplement support from faculty, TAs, and instructors.
 3. **Marks Family Writing Center**: Provides expert help in writing for undergraduate and graduate students.
 4. **Language Direct**: Provides tutoring for foreign languages.
 5. **Van Pelt Library**: Support for students in research and instructional technologies through a range of workshops and consultations.

Canvas

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

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

We will use the Canvas Discussion feature rather than the **Piazza** forum for our class discussions.

Plan of Lecture

- Basic Concepts & Introduction
- Random-Effects & Fixed-Effects Models
- An Introduction to Multilevel Models, I
- Multilevel Models, II
- Multilevel Models, III
- Covariance Pattern Models & GEE

- Mid-Term Examination (Open Book)
- An Introduction to Life Tables and Survival Analysis
- Discrete-Time Duration Models
- Continuous-Time Duration Models
- Cox Regression
- Group-Based Trajectory Models

Class Schedules (Subject to Change)

Week 1 (September 3): *Basic Concepts & Introduction*

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

Week 2 (September 10): *Random-Effects and Fixed-Effects Models*

- Chapter 14 in Wooldridge 2019 (introductory); Chapter 10 in Wooldridge 2010;
- For sociological examples, see Chapter 15 in Treiman
- Western, Bruce. 2002. "The Impact of Incarceration on Wage Mobility and Inequality." *American Sociological Review* 67(4):526-46.
- Laird, N. M. & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, 38: 963-974.

Week 3 (September 17): *An Introduction to Multilevel Models, I*

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

Week 4 (September 24): *Multilevel Models, II*

- Chapter 5 in Singer & Willett.
- Chapters 3, 4, 5 in Raudenbush & Bryk
- Xie, Yu, and Emily Hannum. 1996. "Regional Variation in Earnings Inequality in Reform-Era Urban China." *American Journal of Sociology* 101, no. 4: 950-992.
- 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.
- Song, Xi, and Robert D. Mare. 2019. "Shared Lifetimes, Multigenerational Exposure, and Educational Mobility." *Demography* 56, no. 3: 891-916. (Read the regression analyses on educational mobility)

Week 5 (October 1): *Multilevel Models, III*

- Exercise 2 Due
- Chapter 6 in Singer & Willett.
- Chapters 2, 6, 9 in Raudenbush & Bryk
- Cheng, Siwei. 2014. "A Life Course Trajectory Framework for Understanding the Intracohort Pattern of Wage Inequality." *American Journal of Sociology* 120, no. 3: 633-700.
- Raudenbush, Stephen W., Robert T. Brennan, and Rosalind C. Barnett. 1995. "A Multivariate Hierarchical Model for Studying Psychological Change within Married Couples." *Journal of Family Psychology* 9, no. 2: 161-174.

Week 6 (October 8): *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.

Mid-Term Examination (October 15): *Open Book*

- Exercise 3 Due
- The exam question answers will be discussed in the next week's class.

Week 8 (October 22): *An Introduction to Life Tables and Survival Analysis*

- Chapters 9, 10 in Singer & Willett.
- Chapter 6 in Powers & Xie
- Chapter 3 in Preston, Heuveline, Guillot
- Jenkins, Stephen P. 2005. *Survival Analysis* ch. 1-2 (pp. 1-24). Download [here](#)
- Singer, Judith D., and John B. Willett. 1991. "Modeling the Days of Our Lives: Using Survival Analysis when Designing and Analyzing Studies of Duration and the Timing of Events." *Psychological Bulletin*, Vol. 110(2), pp. 268-290.
- Willett, John B., and Judith D. Singer. 1991. "From Whether to When: New Methods of Studying Student Dropout and Teacher Attrition." *Review of Educational Research*, Vol. 61, pp. 407-450.
- Kaplan, Edward L. and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association*, Vol. 53, pp. 457-481.

Week 9 (October 29): *Discrete-Time Duration Models*

- Chapter 11, 12 in Singer & Willett
- Allison, Paul. 1982. "Discrete-Time Methods for the Analysis of Event Histories." In *Sociological Methodology 1982*, ed. S. Leinhardt, San Francisco, CA: Jossey-Bass, pp. 61-98.
- Raftery, Adrian E., Steven M. Lewis, Akbar Aghajanian, and Michael J. Kahn. 1996. "Event History Modeling of World Fertility Survey Data." *Mathematical Population Studies*, Vol. 6(2), pp. 129-153.
- Raftery, Adrian E., Steven M. Lewis, and Akbar Aghajanian. 1995. "Demand or Ideation? Evidence from the Iranian Marital Fertility Decline." *Demography*, Vol. 32(2), pp. 159-182.
- Lewis, Steven M. and Adrian E. Raftery. 1999. "Bayesian Analysis of Event History Models with Unobserved Heterogeneity via Markov Chain Monte Carlo." *Sociological Methods & Research*, Vol. 28, pp. 35-60.
- Singer, Judith D., and John B. Willett. 1993. "It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events." *Journal of Educational Statistics*, Vol. 18(2), pp. 155-195.
- Jenkins, Stephen P. 1997. "Discrete Time Proportional Hazards Regression." *Stata Technical Bulletin*, STB-39, pp. 22-32.

- Jenkins, Stephen P. 1995. "Easy Estimation Methods for Discrete-Time Duration Models." *Oxford Bulletin of Economics and Statistics*, Vol. 57, pp. 129-138.
- Holford, Theodore R. 1976. "Life Tables with Concomitant Information." *Biometrics*, Vol. 32, pp. 587-597.
- Prentice, R.L. and L.A. Gloeckler. 1978. "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." *Biometrics*, Vol. 34, pp. 57-67.
- Doksum, Kjell A. and Miriam Gasko. 1990. "On a Correspondence between Models in Binary Regression Analysis and in Survival Analysis." *International Statistical Review*, Vol. 58, pp. 243-252.
- Firth, David and Clive Payne, Joan Payne. 1999. "Efficacy of Programmes for the Unemployed: Discrete Time Modelling of Duration Data from a Matched Comparison Study." *Journal of the Royal Statistical Society, Series A*. Vol. 162, pp.111-120.
- Zeng, Zhen, and Yu Xie. 2014. "The Effects of Grandparents on Children's Schooling: Evidence from Rural China." *Demography* 51, no. 2: 599-617.

Week 10 (November 5): Parametric Methods for Continuous-Time Data

- Exercise 4 Due
- Chapter 13 in Singer & Willett
- Jenkins, Stephen P. 2005. *Survival Analysis*. ch.3 (pp. 25-54).
- Bruderl, J. and A. Diekmann. 1995. "The log-logistic Rate Model." *Sociological Methods & Research*, Vol. 24, pp. 158-186.
- Defo, Barthelemy Kuate. 1997. "Effects of Infant Feeding Practices and Birth Spacing on Infant and Child Survival: A Reassessment from Retrospective and Prospective Data." *Journal of Biosocial Science*, Vol. 29, pp. 303-326.
- Defo, Barthelemy Kuate and Alberto Palloni. 1995. "Determinants of Mortality Among Cameroonian Children: Are the Effects of Breastfeeding and Pace of Childbearing Artifacts?" *Genus*, Vol. 51(3-4), pp. 61-96.
- Diekmann, Andreas and Henriette Engelhardt. 1999. "The Social Inheritance of Divorce: Effects of Parent's Family Type in Postwar Germany." *American Sociological Review*, Vol. 64, pp. 783-793.

Week 11 (November 12): Cox Regression

- Chapter 14 in Singer & Willett
- Mare, Robert. D. 1990. "Socio-economic Careers and Differential Mortality Among Older Men in the United States." Pp. 362-87 in Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Comparative Studies of Mortality and Morbidity: Old and New Approaches to Measurement and Analysis*. London: Oxford Univ. Press.

- Piliavin, Irving, and Bradley R. Entner Wright, Robert D. Mare, and Alex H. Westerfelt. 1996. "Exits from and Returns to Homelessness." *Social Service Review*, Vol. 70(1), pp. 33-57.
- Royston, Patrick (2001). "Flexible Parametric Alternatives to the Cox Model, and More." *The Stata Journal*, Vol. 1, pp. 1-28.
- Royston, Patrick and M.K.B. Parmar. (2002). "Flexible Parametric Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects." *Statistics in Medicine*, Vol. 21, pp. 2175-2197.
- Cox, David R. (1972). "Regression Models and Life Tables," *Journal of the Royal Statistical Society, Series B*, Vol. 34, pp. 187-220 (with discussion).
- Menken, Jane, James Trussell, Debra Stempel, and Ozer Babakol. 1981. "Proportional Hazards Life Table Models: An Illustrative Analysis of Socio-Demographic Influences on Marriage Dissolution in the United States." *Demography*, Vol. 18(2), pp. 181-240

Week 12 (November 19): An Introduction to Group-Based Trajectory Models

- Exercise 5 Due
- Presentation by Allison Dunatchik
- Jones, Bobby L. and Daniel S. Nagin. 2007. "Advances in Group-Based Trajectory Modeling and a SAS Procedure for Estimating Them." *Sociological Methods and Research* 35:542-571.
- Nagin, Daniel S. 2005. *Group-Based Modeling of Development*. Boston, MA: Harvard University Press.
- Song, Xi, Emma Zang, and Kenneth C. Land. "Intergenerational Associations of Income Dynamics: A Dyadic Group-Based Approach." working paper.
- Nagin, Daniel S. 1999. "Analyzing Developmental Trajectories: A Semiparametric, Group-based Approach." *Psychological Methods* 4, no. 2: 139-157.
- Laub, John H., Daniel S. Nagin, and Robert J. Sampson. 1998. "Trajectories of Change in Criminal Offending: Good Marriages and the Desistance Process." *American Sociological Review*, 225-238.
- D'unger, Amy V., Kenneth C. Land, Patricia L. McCall, and Daniel S. Nagin. 1998. "How Many Latent Classes of Delinquent/Criminal Careers? Results from Mixed Poisson Regression Analyses." *American Journal of Sociology* 103, no. 6: 1593-1630.
- 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.

Week 13 (November 26): *Thanksgiving, no class*

Week 14 (December 3): *Presenting and Publishing Your Research*

- Recommended Reading: Miller 2005 Chs. 11, 12, 15*;
- Due: First draft of term papers to paired reviewers;
- Assignment: Work on term papers; Reviewer comments for paired term papers;

Week 15 Presenting and Publishing Your Research & Wrap-up (December 10)

- Course summary and discussions.

Advanced Topics

If time permits, we will discuss advanced topics on sequence analysis, multivariate mixed-effects models, and missing data problems in panel data analyses. Students (individuals or groups of students) will discuss a topic in depth for the entire group. The instructor will work with the students in advance of their presentations, reviewing and participating in the formal presentations.