

SOCY 7706: Longitudinal Data Analysis*

Wen Fan

Fall 2025

Term: **Fall 2025**

Class location: **O'Neill 245**

Office location: **McGuinn 408**

Email: **wen.fan@bc.edu**

Instructor: **Dr. Wen Fan**

Class times: **T/Th (10:30-11:45 am)**

Office hours: **Th 3-3:30 pm & by appt**

Mailbox: **McGuinn 426**

1 Course Description

1.1 Overview

This course is designed for graduate students with a prior background in statistics at the level of SOC7703: *Multivariate Statistics* or its equivalent. This means that at the time you enroll in this course, I expect you to have a high degree of competence with basic statistics, including measures of central tendency and dispersion, analysis of variance (ANOVA), and multivariate OLS regression. For those of you in the sociology department, this course provides a foundation for the "Advanced Quantitative Methods" area exam.

This course will introduce several methods for analyzing longitudinal data:

- Fixed-effects models
- Mixed-effects models
- Duration models (also known as event history analysis or survival analysis)
- Sequence analysis
- Time series models

Throughout the course, we will work through empirical pieces by some top social science researchers. As you develop your own research, it is useful to see how others have translated propositions into testable hypotheses, devised appropriate methodologies to test them, and presented the results to diverse audiences. This process will sharpen your written and oral research presentation skills.

1.2 Software

We will be using both Stata and R as the primary software tools. Familiarity with both is required for this course.

*This syllabus was heavily influenced by Christopher Uggen's 2013 syllabus for Event History & Panel Methods at the University of Minnesota

1.3 Course Communication

Please do not hesitate to email me (wen.fan@bc.edu) whenever you have questions. If you don't hear within 24 hours, try again. I am also happy to meet with you at a mutually convenient time.

After class, I will communicate with you by email. Please check your email regularly. If you prefer that I use an email address different from the one you have provided to the university, send me an email from that account indicating your preference.

All course materials will be posted at <http://www.wenfan.co/teaching/longitudinal/>. I will announce the password in our first class.

1.4 Texts and Materials

Recommended:

P.D. Allison (2009). *Fixed Effects Regression Models*. Quantitative Applications in the Social Sciences. SAGE Publications.

P.D. Allison (2014). *Event History and Survival Analysis: Regression for Longitudinal Event Data*. Quantitative Applications in the Social Sciences. SAGE Publications.

Janet M. Box-Steffensmeier et al. (2004). *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press.

S. Rabe-Hesketh and A. Skrondal (2008). *Multilevel and Longitudinal Modeling Using Stata, Second Edition*. Stata Press.

Judith D. Singer and John B. Willett (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press.

Jeffrey M. Wooldridge (2003). *Econometric Analysis of Cross Section and Panel Data*. MIT Press. VII, 211.

2 Course Policies

2.1 Academic Integrity

I have no tolerance for cheating in any form. Such will earn the student an automatic zero. Academic misconduct includes (but is not limited to): cheating on exams; using material from the internet without citing it; plagiarizing any part of work done by someone else; and submitting substantially similar work for two courses without consent. It is your responsibility to familiarize yourself with the university's policy on academic integrity: <https://www.bc.edu/content/bc-web/academics/sites/university-catalog/policies-procedures/academic-integrity.html>. If you have any questions, always consult with me.

2.2 Withdrawals

If you decide to discontinue the course for any reason, please make an official withdrawal. If you fail to officially withdraw from a class which you are no longer attending, you may receive an F on your permanent transcript.

I will not give incompletes.

2.3 Absences

You are expected to come to class prepared to discuss the readings. If you should be absent for unavoidable reasons, you must check with me first.

2.4 Late Work

Up to one week late: 70% of grade awarded (e.g., an A- becomes a C+).

Two weeks late: 35% of grade awarded (e.g., an A- becomes a D).

Later than that: Forget it.

2.5 Classroom Conduct

Please turn off cell phones or other electronic devices that may disrupt class. If you know you need to leave class early, it is less distracting for me if you let me know before class starts. Other disruptive behaviors in class include (but are not limited to): using your laptop for purposes other than note taking, engaging in personal conversations, arriving late, making rude and sarcastic comments.

2.6 Disability Statement

I am committed to supporting the learning of all students in my class. If you have already registered with Connors Family Learning Center (learning disabilities and ADHD) or Disability Services (all other disabilities including temporary disabilities) and have your letter of accommodations, please meet with me early in the course to discuss, plan, and implement your accommodations in the course. If you have or think you have a disability but are not registered, the Connors Family Learning Center and Disability Services websites provide information on the registration process.

3 Assessments and Grades

3.1 One Publishable Paper

The main requirement is **one empirical research paper** that applies some of the methods covered in class to a substantive problem in social science. The paper should include problem formulation, linkage to substantive literature, analysis of suitable data, and discussion of findings. Purely methodological papers will not be accepted. Each of you is starting in a different place, so I expect great variation in the sorts of projects that you produce. Do not waste your time on a paper that will only be used to meet course requirements. The seminar project should advance your own career and research agenda. If you are uncertain about which research question(s) to pursue, please come talk with me as soon as possible. Together, we can develop a project idea that aligns with your research interests and the objectives of this course.

The independent variables (of which there should be at least 5) must include at least one continuous measure and one categorical measure. If your data include a sample weight variable, you should weight all your analysis. If there are multiple sample weights (as is often the case), consult with me if you do not know which one to use.

I highly recommend you submit a two-page proposal that outlines the problem, data, and empirical analyses. The proposal is optional, but I promise feedback on proposals that come in by **October 26**. This proposal should consist of the following elements: (1) introduction to the problem, (2) data and methods to use, and (3) expected analyses.

In our last class (**December 4**), every student will give a presentation, showing (preliminary) results. These are not a big part of your grade, but in my opinion are truly essential to your professional success.

The final paper should be no more than 25 pages (double-spaced, excluding references, tables, and figures). It must come with an Appendix which includes all codes used in the analysis of this project. The paper is due on **December 12**.

3.2 Assignments

There will be **three homework assignments**. To obtain credit for homework, you must hand it in on time and be prepared to present solutions orally in class. Homework assignments are meant to achieve three goals: (1) to provide practice with Stata/R, (2) to provide practice with the statistical concepts discussed in class, and (3) to provide a chance to demonstrate your mastery of material and highlight areas where more work is needed. You may work in a group consisting of up to two persons in size, but all write-ups must be done independently. All collaborators should be appropriately cited in your write-up.

3.3 Readings and AI Presentation

Weekly readings are assigned on the course syllabus. All readings are assumed to be completed *before* each lecture.

You will deliver an in-class **presentation** reflecting on your analytical journey, comparing AI responses with the methods and interpretations in a published sociological paper using longitudinal data. Begin by explaining the dataset's structure, key variables, and research questions to an AI language model (e.g., ChatGPT, Claude). Document the AI's suggested analytical methods, assumptions, and interpretations. In your presentation, evaluate the AI's output and contrast it with the methodology used in the original research paper, highlighting any methodological errors, omissions, or unrealistic assumptions. You are encouraged to reflect on the strengths and limitations of both the AI-generated and human-authored approaches. We will discuss further details about this activity later in the course.

3.4 Participation and Collaboration

Build class camaraderie: prepare, participate, help others. Seminars are constructed in interaction. I will try to provide an environment in which everyone feels comfortable participating, but responsibility for the seminar is borne collectively. This means that you must come to class prepared to discuss the readings (preferably with written comments and at least one question of your own) and having considered how the course materials will affect your work.

3.5 Grading

The term paper determines 50 percent of the final grade. Homework assignments, presentation, and class participation determine 30, 10, and 10 percent, respectively.

Letter grade assignment:

A	93%+
A-	90-92.99%
B+	87-89.99%
B	83-86.99%
B-	80-82.99%
C+	77-79.99%
C	73-76.99%
C-	70-72.99%
D	60-69.99%
F	<60%

4 Data

You are welcome to work with any dataset you choose. Some commonly used social science datasets are listed below. If you plan to use a dataset that is not publicly available, it is especially important to get started early so you have enough time to request and receive the data. Some of the datasets listed below are available through [ICPSR](#), rather than directly from the study website. The [Atlas of Longitudinal Datasets](#) is also a great resource for identifying potential longitudinal datasets.

British Household Panel Survey: <https://www.iser.essex.ac.uk/bhps/acquiring-the-data>
Current Population Survey: <http://www.census.gov/cps>
Early Childhood Longitudinal Study: <http://nces.ed.gov/ecls/index.asp>
Future of Families & Child Wellbeing Study: <https://ffcws.princeton.edu>
General Social Survey: <http://www3.norc.ox.ac.uk/gss+website>
German Socio-Economic Panel Study: <http://www.diw.de/en/soep>
Health and Retirement Study: <http://hrsonline.isr.umich.edu>
High School & Beyond: <http://nces.ed.gov/surveys/hsb>
International Social Survey Programme: <http://www.issp.org/index.php>
National Longitudinal Study of Adolescent Health: <https://addhealth.cpc.unc.edu>
National Longitudinal Surveys: <http://www.bls.gov/nls>
National Social Life, Health, and Aging Project: <https://www.norc.ox.ac.uk/research/projects/national-social-life-health-and-aging-project.html>
Panel Study of Income Dynamics: <https://psidonline.isr.umich.edu/default.aspx>
Survey of Income and Program Participation: <https://www.census.gov/programs-surveys/sipp.html>
Wisconsin Longitudinal Study: <https://wls.wisc.edu>

5 Readings

August 26 and 28: Welcome and theoretical underpinnings

- **Lecture:** introduction; course logistics; requirements; theoretical underpinnings of studying time
- **Readings:**
 - Glen H. Elder, Monica Kirkpatrick Johnson, and Robert Crosnoe (2003). “The Emergence and Development of Life Course Theory”. In: *Handbook of the Life Course*. Ed. by Jeylan T. Mortimer and Michael J. Shanahan. Red. by Howard B. Kaplan. Boston, MA: Springer US, pp. 3–19. doi: [10.1007/978-0-306-48247-2_1](https://doi.org/10.1007/978-0-306-48247-2_1)
 - Shoshana R. Dobrow and Hannah Weisman (2021). “Only Time Will Tell: Conducting Longitudinal Research on Careers”. In: *Handbook of Research Methods in Careers*. Ed. by Wendy Murphy and Jennifer Tosti-Kharas. Edward Elgar Publishing. doi: [10.4337/9781788976725.00016](https://doi.org/10.4337/9781788976725.00016)

September 2 and 4: Why OLS models are not sufficient?

- **Lecture:** OLS assumptions; lagged dependent variables; change score
 - **Readings:** CT 3.4-3.5.12
-

September 16 and 18: Fixed-effects models

- **Lecture:** unobserved heterogeneity; least squares dummy variable; demeaning; difference-in-differences (two-way fixed effects, TWFE); hybrid models
 - **Readings:**
 - [Chapter 16 - Fixed Effects](#) (skip the two subsections on random effects)
 - [Chapter 18 - Difference-in-Differences](#)
 - Trenton D. Mize and Reilly Kincaid (2025). “Role-Accumulation and Mental Health across the Life Course”. In: *American Sociological Review* 90.2, pp. 226–256. doi: [10.1177/00031224241313394](https://doi.org/10.1177/00031224241313394)
-

September 23 and 25: Mixed-effects models

- **Lecture:** random effects; random intercepts models; random slopes models; variance components; intra-class correlation (ICC); centering; growth curve models
 - **Reading:** Markus Brauer and John J. Curtin (2018). “Linear Mixed-Effects Models and the Analysis of Nonindependent Data: A Unified Framework to Analyze Categorical and Continuous Independent Variables That Vary within-Subjects and/or within-Items.” In: *Psychological Methods* 23.3, pp. 389–411. doi: [10.1037/met0000159](https://doi.org/10.1037/met0000159)
-

September 30 and October 2: Mixed-effects models

- **Lecture:** estimation methods; cross-level interactions; model fit; model diagnostics
 - **Reading:** Mengjie Cheng et al. (2023). “How New Ideas Diffuse in Science”. In: *American Sociological Review* 88.3, pp. 522–561. doi: [10.1177/00031224231166955](https://doi.org/10.1177/00031224231166955)
-

October 7 and 9: Mixed-effects models

- **Lecture:** crossed versus nested random effects; three-level models; generalized mixed models; Bayesian multilevel models
 - **Reading:** Paul-Christian Bürkner (2017). “Brms: An R Package for Bayesian Multilevel Models Using Stan”. In: *Journal of Statistical Software* 80, pp. 1–28. doi: [10.18637/jss.v080.i01](https://doi.org/10.18637/jss.v080.i01)
-

October 16: Duration models

- **Lecture:** event; censoring; truncation; survival function; hazard function; expectation of life; median survival time; life table estimator; Kaplan-Meier estimator
 - **Readings:** [Event History Analysis](#) (pages 369–373)
-

October 21 and 23: Duration models

- **Lecture:** parametric models; accelerated failure time; proportional hazards; exponential; Weibull; Gompertz; log-normal; log-logistic; generalized Gamma; semi-parametric models; Cox; proportional hazards assumption
 - **Readings:**
 - [Event History Analysis](#) (pages 373–378)
 - Susan Olzak (2021). “Does Protest Against Police Violence Matter? Evidence from U.S. Cities, 1990 through 2019”. In: *American Sociological Review* 86.6, pp. 1066–1099. DOI: [10.1177/00031224211056966](#)
-

October 28 and 30: Duration models

- **Lecture:** discrete-time; complementary log-log; repeated events; competing risks
 - **Readings:**
 - [Event History Analysis](#) (pages 379–384)
 - Louise Caron, Haley McAvay, and Mirna Safi (2023). “Born Again French: Explaining Inconsistency in Citizenship Declarations in French Longitudinal Data”. In: *American Sociological Review* 88.6, pp. 1066–1103. DOI: [10.1177/00031224231207392](#)
-

November 4 and 6: Sequence analysis

- **Lecture:** dissimilarity; optimal matching; insertion/deletion; substitution; steps of sequence analysis; cluster analysis
 - **Readings:**
 - [TraMineR at a glance](#)
 - Anette Eva Fasang et al. (2024). “Lives in Welfare States: Life Courses, Earnings Accumulation, and Relative Living Standards in Five European Countries”. In: *American Journal of Sociology* 130.2, pp. 384–438. DOI: [10.1086/730851](#)
-

November 11 and 13: Markov models

- **Lecture:** Markov chains; transition modeling; hidden Markov models (HMMs)
 - **Reading:** Satu Helske and Jouni Helske (2019). “Mixture Hidden Markov Models for Sequence Data: The seqHMM Package in R”. in: *Journal of Statistical Software* 88, pp. 1–32. DOI: [10.18637/jss.v088.i03](#)
-

November 20 and December 2: Time series models

- **Lecture:** time series; autoregression; non-stationarity; cointegration; Nickell bias
 - **Readings:** Adaner Usmani (2018). "Democracy and the Class Struggle". In: *American Journal of Sociology* 124.3, pp. 664–704. DOI: [10.1086/700235](https://doi.org/10.1086/700235)
-

December 4: Student presentations