

SOCY 7717: Event History Analysis and Sequence Analysis *

Wen Fan

Spring 2019

Term: **Spring 2019**

Class location: **O'Neill 245**

Office location: **McGuinn 408**

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

Instructor: **Dr. Wen Fan**

Class times: **TuTh (10:30-11:45 am)**

Office hours: **Tu (12 - 1 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 *SOC7704: Regression Models for Categorical Data* 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 intermediate statistics, including analysis of variance (ANOVA), multivariate OLS regression, and binary logit regression. It would be ideal for students considering a dissertation using event history analysis or sequence analysis, but also appropriate for those who just want more hands-on experience with methodologies that have become increasingly popular in both scientific and policy work. For those of you in the sociology department, this course provides a foundation for the "Advanced Quantitative Methods" area exam.

Event history analysis can be used when researchers want to predict *whether and when* something happens—for example, wars, births, deaths, strikes, crimes, or job promotions. Using examples that take nations, states, and individuals as the units of analysis, the first half of the course will cover topics such as demographic life tables, survival and hazard analysis, competing risks, proportional hazards, and time-varying covariates. In the second half of the course, we will focus on sequence analysis, which is ideal when capturing the *holistic sequence* is the main interest. Throughout the course we will focus on the basics and learn to apply these basics well.

I have several objectives for this class.

- First, I will help you achieve a deeper and more nuanced understanding of two important classes of quantitative research techniques in the social sciences.
- Second, this course will provide support, encouragement, and an opportunity for you to make significant progress on one or more of your own research projects.
- Third, in using and explaining these models, you will sharpen your written and oral research presentation skills.

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

- Fourth, we will work through empirical pieces by some top 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.

1.2 Software

All of the models discussed in this class can be estimated in Stata or R. Familiarity with Stata or R is not required for this course.

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. You can also visit me during my office hour. If the current office hour does not work for you, I am happy to meet with you at another mutually convenient time. 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/ehasa/>. I will announce the password during our first class.

1.4 Texts and Materials

Required:

Cleves, Mario, William W. Gould, and Yulia V. Marchenko. 2016. *An Introduction to Survival Analysis Using Stata, Revised Third Edition*. College Station, TX: Stata Press. [EH]

We will also read some challenging research/methodology articles during the semester, but I've limited the number of required readings to just a few—emphasizing solid and accessible applications of the techniques we discuss, rather than the brilliant statistical innovations that brought us these techniques.

Recommended:

Cornwell, Benjamin. 2015. *Social Sequence Analysis: Methods and Applications*. Cambridge, MA: Cambridge University Press.

Box-Steffensmeier, J.M. and Jones, B.S. 2004. *Event History Modeling: a Guide for Social Scientists*. Cambridge University Press, MA: Cambridge.

Singer JD, Willett JB. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.

Long, J. Scott. 2008. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press. (If you plan to do a lot of data analysis, this book will save you tons of time and make your work replicable.)

Gandrud, Christopher. 2015. *Reproducible Research with R and RStudio, Second Edition*. Boca Raton, FL: CRC Press Book. (Similar as Long's book, but with a target audience of R users.)

1.5 Course Feedback

Any feedback is greatly appreciated. In particular, I will conduct a few course evaluations in the end of the Thursday class throughout the semester. I will hand over a review form to you consisting of 3-4 short questions, such as whether or not the articles you read were helpful, whether or not a particular technique we covered was easy to understand, anything you might need further explanation, etc. This review is conducted for my own improvement and for the benefit of future students; it is anonymous and won't be used against you.

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/offices/stserv/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, and then choose a seat close to the exit. 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

Boston College is committed to providing reasonable accommodations and integrated access for students with disabilities to all available academic, social, and recreational programs and activities. Appropriate support and referral services are provided by the Disability Services Office, which serves students with hearing, visual, mobility, medical, and psychiatric disabilities. If you are a student with a documented disability seeking reasonable accommodations in this course, please contact Kathy Duggan, (617) 552-8093, dugganka@bc.edu, at the Connors Family Learning Center regarding learning disabilities and ADHD, or Paulette Durrett, (617) 552-3470, paulette.durrett@bc.edu, in the Disability Services Office regarding all other types of disabilities, including temporary disabilities. Advance notice and appropriate documentation are required for accommodations. If you are unsure whether or not Disability Services are appropriate for your needs, please make an appointment with Disability Services for a consultation.

3 Assessments and Grades

3.1 One Publishable Paper

The main requirement is **one empirical research paper** that applies either event history analysis or sequence analysis 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. Every student 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.

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 **March 2**. This proposal should consist of the following elements: (1) introduction to the problem, (2) hypotheses to be tested, (3) methods to use, and (4) expected results.

In our last class (**May 2**), every student will give a presentation, showing (preliminary) results. These are not a big part of your grade, but in my opinion truly essential to your professional success.

The final paper should be no more than 25 pages (excluding references, tables, and figures). It must come with an Appendix which includes all Stata or R codes (or codes from whatever software you use) used in the analysis of this project. The paper is due on **May 7**.

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 or 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; any Stata/R code from a given source should also be cited as you would a journal article.

3.3 Readings and Case Studies

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

Case studies are empirical pieces to help you understand how to apply the methods we learn in class. They are required readings, which we will examine closely in class (**March 21** and **April 30**). The class will be divided into pairs to play the roles of "reviewers" and "authors". For the provided paper, "reviewers" are expected to comment on its contribution, as well as to offer critiques (with a focus on methodologies or methods-related substantive issues). "Authors" are expected to defend themselves (i.e., are reviewers' critiques reasonable?); if critiques are valid, "authors" need to indicate the types of analyses they would perform to address them. Whatever role you are assigned, please be brief and to-the-point in your write-up and in-class presentation. More details will be discussed at a later point in class.

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 to have considered how the course materials will affect your work. A rough guideline: speak at least once during each course meeting.

3.5 Grading

The term paper determines 50 percent of the final grade. Homework assignments, case studies, 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 wish. Some commonly-used longitudinal datasets are listed below. If you wish to use a dataset that is not publicly available, it is especially important that you get started quickly in order to have time to request and receive the data. Some datasets are available through **ICPSR**, rather than directly from the study website.

British Household Panel Survey: <https://www.iser.essex.ac.uk/bhps>

Cape Area Panel Study: <http://www.caps.uct.ac.za>

Current Population Survey: <http://www.census.gov/cps>

Early Childhood Longitudinal Study: <http://nces.ed.gov/ecls/index.asp>

Fragile Families and Child Wellbeing Study: <http://www.fragilefamilies.princeton.edu>

General Social Survey: <http://www3.norc.org/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>

Los Angeles Family and Neighborhood Survey: <http://lasurvey.rand.org>

National Longitudinal Study of Adolescent Health: <http://www.cpc.unc.edu/projects/addhealth>

National Longitudinal Surveys: <http://www.bls.gov/nls>

National Social Life, Health, and Aging Project: <http://www.norc.org/Research/Projects/Pages/national-social-life-health-and-aging-project.aspx>

Panel Study of Income Dynamics: <http://www.psidonline.isr.umich.edu>

Project on Human Development in Chicago Neighborhoods: <http://www.icpsr.umich.edu/icpsrweb/PHDCN/about.jsp>

Survey of Income and Program Participation: <http://www.census.gov/sipp>

Wisconsin Longitudinal Study: <http://www.ssc.wisc.edu/wlsresearch>

5 Class Schedule Overview

The following schedule is tentative. Please keep yourself updated by checking email regularly.

	Week	Lecture	Due
I	1/15	Introduction (course, Stata)	
	1/22	Event history basics	
	1/29	Nonparametric methods	
	2/5	Prepare survival data for analysis	
	2/12	Parametric methods	
	2/19	Parametric/Semiparametric methods	Assignment 1 (2/19)
	2/26	Semiparametric methods	Proposal (optional, 3/2)
	3/12	Discrete-time event history analysis	
	3/19	Advanced topics in event history analysis	Case study 1 (3/21)
II	3/26	Introduction to R	Assignment 2 (3/26)
	4/2	Describing and comparing state sequences	
	4/16	Sequence analysis	
	4/23	Advanced topics in sequence analysis	Assignment 3 (4/23)
	4/30	Review and presentations	Case study 2 (4/30) and project (5/7)

6 Readings

January 15: Welcome and math review

- **Lecture:** introduction; course logistics; requirements; math review

January 17: Introduction to Stata

- **Lecture:** basic Stata commands
-

January 22: Event history basics I

- **Lecture:** event; censoring; truncation
- **Readings:** EH Chapters 1 and 4

January 24: Event history basics II

- **Lecture:** survival function; hazard function; ratio vs. proportion vs. rate; expectation of life; median survival time
 - **Readings:** EH Chapters 2 and 3 (skip 2.2)
-

January 29: Nonparametric methods I

- **Lecture:** life table estimator; Kaplan-Meier estimator
- **Readings:** begin EH Chapter 8

January 31: Nonparametric methods II

- **Lecture:** Nelson-Aalen estimator; log-rank test; Wilcoxon test
 - **Readings:** EH Chapter 8
-

February 5: Prepare survival data for analysis

- **Lecture:** `stset`; `stsum`; `stdes`; `stptime`; `strate`
- **Readings:** EH Chapter 5; begin EH Chapter 6

February 7: Prepare survival data for analysis

- **Lecture:** `ltable`; `sts list`; `sts graph`; `stci`; `sts test`
 - **Readings:** EH Chapters 6-7
-

February 12: Parametric methods I

- **Lecture:** parametric models; accelerated failure time; proportional hazards; exponential; Weibull; Gompertz; log-normal; log-logistic; generalized Gamma
- **Readings:** EH Chapters 12-13

February 14: Parametric methods II

- **Lecture:** `streg`; MLE
 - **Readings:** EH Chapter 14
-

February 19: Parametric methods III

- **Lecture:** `stcurve`; marginal effects; model comparisons; Cox-Snell residuals
- **Readings:** EH Chapter 14

February 21: Semiparametric methods I

- **Lecture:** semi-parametric models; Cox; partial likelihood; `stcox`
 - **Readings:** EH Chapter 9 (skip 9.3 and 9.4)
-

February 26: Semiparametric methods II

- **Lecture:** proportional hazards assumption; Schoenfeld residuals; Martingale residuals; `estat phtest`; `stphplot`
- **Readings:** EH Chapter 11

February 28: Semiparametric methods III

- **Lecture:** time-varying variables; `stsplit`; stratification
 - **Readings:** EH 9.3 and Chapter 10
-

March 12: Discrete-time event history analysis

- **Lecture:** discrete-time; `stsplit`; complementary log-log; Poisson
- **Readings:** re-read Chapter 6 if needed

March 14: Repeated events and competing risks

- **Lecture:** repeated events; marginal methods; conditional methods; parallel processes; competing risks; conditional processes
 - **Readings:** EH Chapter 17
-

March 19: Frailty models

- **Lecture:** heterogeneity; frailty models
- **Readings:** EH Chapter 15

March 21: Case study and review

- **Lecture:** review
 - **Case study:** Warner, Tara D., David F. Warner, and Danielle C. Kuhl. 2017. "Cut to the Quick: The Consequences of Youth Violent Victimization for the Timing of Dating Debut and First Union Formation." *American Sociological Review* 82(6): 1241-1271.
-

March 26: Introduction to R (I)

- **Lecture:** basic R commands

March 28: Introduction to R (II)

- **Lecture:** basic R commands
-

April 2: Describe/Visualize state sequences

- **Lecture:** sequence analysis; life course; timing; state sequences; holistic; TraMineR; sequence index plots; sequence frequency table; transition rates; entropy index; state distribution plots; complexity index; seqIplot; seqtab; seqmtplot; seqtrate; seqstatd; seqdplot; seqHtplot; seqmsplot; seqtransn; seqient
- **Readings:**
 - Gauthier, Jacques-Antoine, Felix Bühlmann, and Philippe Blanchard. 2014. “Introduction: Sequence Analysis in 2014.” In *Advances in Sequence Analysis: Theory, Method, Applications*, pp. 1-17. Springer.
 - Chapters 6 and 7 of the TraMineR manual (<http://mephisto.unige.ch/pub/TraMineR/doc/TraMineR-Users-Guide.pdf>)

April 4: Sequence comparison algorithms

- **Lecture:** dissimilarity; common attributes; edit distances; optimal matching; insertion/deletion; substitution
 - **Readings:** Elzinga, Cees H. 2014. “Distance, Similarity and Sequence Comparison.” In *Advances in Sequence Analysis: Theory, Method, Applications*, pp. 51-73. Springer.
-

April 16: Sequence analysis

- **Lecture:** steps of sequence analysis; cluster analysis
 - **Readings:**
 - The WeightedCluster library manual (<https://cran.r-project.org/web/packages/WeightedCluster/vignettes/WeightedCluster.pdf>)
 - 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.
-

April 23: Network methods for sequence analysis

- **Lecture:** networks of sequences; network concepts and terms; sequence-network construction; event structure analysis
- **Readings:** Bison, Ivano. 2014. “Sequence as Network: An Attempt to Apply Network Analysis to Sequence Analysis.” In *Advances in Sequence Analysis: Theory, Method, Applications*, pp. 231-248. Springer.

April 25: Multichannel sequence analysis and other new developments

- **Lecture:** multichannel sequence analysis; linked lives
 - **Readings:** Gauthier, Jacques-Antoine, Eric D. Widmer, Philipp Bucher, and Cédric Notredame. 2010. “1. Multichannel Sequence Analysis Applied to Social Science Data.” *Sociological Methodology* 40(1): 1-38.
-

April 30: Case study and review

- **Lecture:** review
- **Case study:** Aisenbrey, Silke, and Anette Fasang. 2017. "The Interplay of Work and Family Trajectories Over the Life Course: Germany and the United States in Comparison." *American Journal of Sociology* 122(5): 1448-1484.

May 2: Student presentations
