GSERM - St. Gallen 2019

Longitudinal Data Analysis

Professor Christopher Zorn
Pennsylvania State University
zorn@psu.edu

Course Content

The subject matter of the course is regression models for data that vary both over cross-sectional units and across time. The course will begin with a discussion of the relevant dimensions of variation in such data, and discuss some of the challenges and opportunities that such data provide. It will move on to models for one-way unit effects (fixed, between, and random), models for complex panel error structures, dynamic panel models, and nonlinear models for discrete dependent variables. The second part of the course will focus on models for time-to-event ("survival," or "event history") data. In every case, students will learn the statistical theory behind the various models, details about estimation and inference, and techniques for the substantive interpretation of statistical results. Students will also develop statistical software skills for fitting and interpreting the models in question, and will use the models in both simulated and real data applications. Students will leave the course with a thorough understanding of both the theoretical and practical aspects of conducting analyses of longitudinal data.

Prerequisites (knowledge of topic)

Students should have a comfortable familiarity with univariate differential and integral calculus, basic probability theory, and linear algebra is required. Students should have completed Ph.D.-level courses in introductory statistics and linear regression models, up to the level of Regression III. Familiarity with discrete and continuous univariate probability distributions will be helpful.

Hardware

Course exercises will be completed on the students' own laptop computers. For purposes of this class, a laptop running any widely-used operating system (Windows, OS-X, Linux) will be acceptable.

Software

All lecture materials, slides, and in-class examples will be conducted using the R statistical language. Students are encouraged to come to class with current versions of both R (via `https://www.r-project.org`) and RStudio (`https://www.rstudio.com`) on their laptops. The instructor can also provide limited support for students wishing to use Stata (`http://www.stata.com`). Students electing to use other statistical software (e.g., SAS, PSPP, etc.) may find themselves to be at a substantial disadvantage.

Structure

Day One:

- Morning: Overview of Panel/TSCS data + One-Way Unit Effects
- Afternoon: GLS-ARMA and Dynamic Panel Data Models

Day Two:

- Morning: Hierarchical / Multilevel Models for TSCS Data
- Afternoon: Models for Binary and Event Count Dependent Variables

Day Three:

- Morning: Generalized Estimating Equations
- Afternoon: Introduction to Survival / Event History Data

Day Four:

- Morning: Parametric and Semiparametric Models for Survival Data
- Afternoon: Discrete-Time Models

Day Five:

- Morning: Survival Model Extensions
- Afternoon: Examination

Literature

Mandatory

The course has two required texts:

- Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists.* New York: Cambridge University Press.
- Hsaio, Cheng. 2003. *Analysis of Panel Data.* New York: Cambridge University Press.

Additional readings will also be assigned as necessary, all of which will be available on the course github repository and/or through JSTOR.

Supplementary / Voluntary

None.

Mandatory Readings Before Course Start

None.

Examination

Students will be evaluated on two written homework assignments that will be completed during the course (20 percent each) and a final examination (60 percent). Homework assignments will typically involve a combination of simulation-based exercises and "real data" analyses, and will be completed during the evenings while the class is in session. For the final examination, students will have two alternatives:

1. **"In-Class"**: Complete the final examination in the afternoon of the last day of class (from roughly noon until 6:00 p.m. local time), or
2. **"Take-Home"**: Complete the final examination during the week following the end of the course (due date: 28 June 2019).

Additional details about the final examination will be discussed in the morning session on the first day of the course.

Supplementary aids

The exam will be a "practical examination" (see below for content). Students will be allowed access to (and encouraged to reference) all course materials, notes, help files, and other documentation in completing their exam.

Examination Content

The final examination will involve the application of the techniques taught in the class to one or more "live" data example(s). These will typically take the form of either (a) a replication and extension of an existing published work, or (b) an original analysis of observational data with a survival / duration component. Students will be required to specify, estimate, and interpret various forms of survival models, to conduct and present diagnostics and robustness checks, and to give detailed justifications for their choices.

Literature

Panel / Time-Series Cross-Sectional Models:

- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What To Do (And Not To Do) With Time- Series Cross-Section Data." *American Political Science Review* 89(September): 634-647.
- Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data.* New York: Cambridge University Press. Chapter 9.
- Clark, Tom S. and Drew A. Linzer. 2015. "Should I Use Fixed Or Random Effects?" *Political Science Research and Methods* 3(2):399-408.
- Keele, Luke, and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14(2):186-205.
- Zorn, Christopher. 2001. "Estimating Between- and Within-Cluster Covariate Effects, with an Application to Models of International Disputes." *International Interactions* 27(4):433-45.

- Zorn, Christopher. 2001. "Generalized Estimating Equation Models for Correlated Data: A Review with Applications." *American Journal of Political Science* 45(April):470-90.

Survival / Event History Models:

- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(October):1260-88 (and erratum).
- Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. New York: Cambridge University Press.
- Box-Steffensmeier, Janet M., and Christopher Zorn. 2001. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science* 45(October):951- 67.
- Box-Steffensmeier, Janet M., and Christopher Zorn. 2002. "Duration Models for Repeated Events." *Journal of Politics* 46(November):1069-94.
- Pintilie, Melania. 2007. "Analyzing and Interpreting Competing Risk Data." *Statistics in Medicine* 26:1360-67.
- Signorino, Curt, and David Carter. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3):271-292. Also read response by Beck and rejoinder by Signorino & Carter.
- Zorn, Christopher. 2000. "Modeling Duration Dependence." *Political Analysis* 8(Autumn): 367-380.

Workload

At least 24 units, 45 minutes each, on 5 consecutive days.