GSERM - St. Gallen 2018

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.) will be at a substantial disadvantage.

Structure

Day One:

- **Morning: Overview of Panel/TSCS data + One-Way Unit Effects**

- Hsaio, Cheng. 2003. *Analysis of Panel Data*. Chapters 1 and 3.
- Zorn, Christopher. 2001. "Estimating Between- and Within-Cluster Covariate Effects, with an Application to Models of International Disputes." *International Interactions* 27(4):433-45.

Additional resources:

- Bartels, Larry M. 1996. "Pooling Disparate Observations." *American Journal of Political Science* 40(August):905-42.
- Clark, Tom S. and Drew A. Linzer. 2015. "Should I Use Fixed Or Random Effects?" *Political Science Research and Methods* 3(2):399-408.
- Finkel, Steven E., and Edward N. Muller. 1998. "Rational Choice and the Dynamics of Political Action: Evaluating Alternative Models with Panel Data." *American Political Science Review* 92(March):37-50.
- Neuhaus, J. M., and J. D. Kalbfleisch. 1998. "Between- and Within-Cluster Covariate Effects in the Analysis of Clustered Data." *Biometrics* 54:638-45.
- Nuamah, Nicholas N. N. N. 1986. "Pooling Cross Section and Time Series Data." *The Statistician* 35:345-51.
- Plumper, Thomas, and Vera E. Troeger. 2007. "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects." *Political Analysis* 15(2):124-139.
- Stimson, James. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29:914-47.
- "Symposium on Fixed-Effects Vector Decomposition." 2011. *Political Analysis* 19(2).

- **Afternoon: GLS-ARMA and Dynamic Panel Data 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.
  - Beck, Nathaniel, and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series Cross-Section Models." *Political Analysis* 6:1-36.

Additional resources:

- Achen, Christopher. 2000. "Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables." Presented at the Annual Meeting of the Society for Political Methodology, UCLA. Available here.
- Anderson, T.W., and C. Hsiao. 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." *Journal of Econometrics* 18:47-82.
- Beck, Nathaniel. 1991. "Comparing Dynamic Specifications: The Case of Presidential Approval." *Political Analysis* 3:51-87.
- Beck, Nathaniel. 2001. "Time-Series Cross-Section Data: What Have We Learned in the Past Few Years?" *Annual Review of Political Science* 4:271-293.
- Beck, Nathaniel, and Jonathan Katz. 2011. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." *Annual Review of Political Science* 14:331-52.

- Blais, Andre, Donald Blake, and Stephane Dion. 1996. "Do Parties Make a Difference: A Reappraisal." *American Journal of Political Science* 40:514-520.
- Burkhart, Ross E., and Michael S. Lewis-Beck. 1994. "Comparative Democracy: The Economic Development Thesis." *American Political Science Review* 88:903-910.
- 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.
- Smith, Mark A. 2001. "The Contingent Effects of Ballot Initiatives and Candidate Races on Turnout." *American Journal of Political Science* 45(3): 700-706.
- Wawro, Gregory. 2002. "Estimating Dynamic Panel Data Models in Political Science." *Political Analysis* 10(Winter):25-48.
- Wawro, Gregory, and Ida Pagter Kristensen. 2006. "Lagging the Dog?: The Robustness of Panel Corrected Standard Errors in the Presence of Serial Correlation and Observation Specific Effects." Working paper: Columbia University. Contact Dr. Wawro (`gjw10@columbia.edu`) if you're interested in this paper.
- Wilson, Sven E., and Daniel M. Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis* 15(2):101-123.

Day Two:

- **Morning: Hierarchical / Multilevel Models for TSCS Data**
  - Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press, Chapters 11-12.
  - White, Halbert (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48: 817-838.

  Additional resources:

  - Freedman, D. A. 2006. "On the So-Called 'Huber Sandwich Estimator' and 'Robust' Standard Errors." *The American Statistician* 60:299-302.
  - King, Gary, and Margaret E. Roberts. 2014. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What To Do About It." *Political Analysis* 22:1-21.
  - A million textbooks on multilevel models...

- **Afternoon: Models for Binary and Event Count Dependent Variables**
  - 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.
  - Hsaio, Cheng. 2003. *Analysis of Panel Data*. Chapter 7, §7.1-7.3 and Chapter 8.

  Additional resources:

  - Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press. Chapter 9.

- Green, Donald P., Soo Yeon Kim, and David Yoon. 2001. "Dirty Pool." *International Organization* 55:441-68 (and commentary by Beck & Katz, Oneal & Russett, and King).
- Katz, Ethan. 2001. "Bias in Conditional and Unconditional Fixed Effects Logit Estimation." *Political Analysis* 9(Autumn):379-84 (and also see Coup'e, Tom (2005) "Bias in Conditional and Unconditional Fixed Effects Logit Estimation: A Correction." *Political Analysis* 13(Summer):292-95).
- Li, Quan, and Drew Schaub. 2004. "Economic Globalization and Transnational Terrorism: A Pooled Time-Series Analysis." *Journal of Conflict Resolution* 48:230-258.
- Martin, Andrew D. 2003. "Bayesian Inference for Heterogeneous Event Counts." *Sociological Methods and Research* 32:30-63.
- Wawro, Gregory. 2001. "A Panel Probit Analysis of Campaign Contributions and Roll Call Votes." *American Journal of Political Science* 45(July):563-579.
- Whitford, Andrew B., Jeff Yates, and Holona L. Ochs. 2006. " Ideological Extremism and Public Participation." *Social Science Quarterly* 87(1):36-54.
- Wooldridge, Jeffrey. 1999. "Distribution-Free Estimation of Some Nonlinear Panel Data Models." *Journal of Econometrics* 90(May):77-97.

Day Three:

- **Morning: Generalized Estimating Equations**
  - Zorn, Christopher. 2001. "Generalized Estimating Equation Models for Correlated Data: A Review with Applications." *American Journal of Political Science* 45(April):470-90.
  - Neuhaus, J. M., J. D. Kalbfleisch, and W. W. Hauck. 1991. "A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data." *International Statistical Review* 59(1):25-35.

  Additional resources:

  - Baker, Andy, and Kenneth F. Greene. 2011. "The Latin American Left's Mandate: Free-Market Policies and Issue Voting in New Democracies." *World Politics* 63(1):43-77.
  - Ballinger, Gary A. 2004. "Using Generalized Estimating Equations for Longitudinal Data Analysis." *Organizational Research Methods* 7:127-50.
  - Caldeira, Gregory A., John R. Wright, and Christopher Zorn. 1999. "Strategic Voting and Gatekeeping in the Supreme Court." *Journal of Law, Economics and Organization* 15(3):549-72.
- **Afternoon: Introduction to Survival / Event History Data**
  - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 1-2.

  Additional resources:

- Cioffi-Revilla, Claudio. 1984. "The Political Reliability of Italian Governments: An Exponential Survival Model." *American Political Science Review* 78(2):318-37.
- Hosmer, David W., and Stanley Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*, pp. 27-84 and Appendix 1.
- Zelditch, Morris Jr. and Joan Butler Ford. 1994. "Uncertainty, Potential Power, and Nondecisions." *Social Psychology Quarterly* 57(1):64-73.

Day Four:

- **Morning: Parametric and Semiparametric Models for Survival Data**
  - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapter 3.

  Additional resources:

  - Alt, James, and Gary King. 1994. "Transfers of Governmental Power: The Meaning of Time Dependence." *Comparative Political Studies* 27(2):190-210.
  - Bennett, D. Scott, and Allan C. Stam III. 1996. "The Duration of Interstate Wars." *American Political Science Review* 90(June):239-57.
  - Bueno de Mesquita, Bruce, and Randolph M. Siverson. 1995. "War and the Survival of Political Leaders: A Comparative Study of Regime Types and Political Accountability." *American Political Science Review* 89(2):841-55.
  - Teachman, Jay D., and Mark D. Hayward. 1993. "Interpreting Hazard Rate Models." *Sociological Methods and Research* 21(February):340-71.

- **Afternoon: Cox and Discrete-Time Models**
  - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 4-5.

  Additional resources:

  - Alt, James E., Gary King and Curtis S. Signorino. 2001. "Aggregation Among Binary, Count and Duration Models: Estimating the Same Quantities from Different Levels of Data." *Political Analysis* 9(Winter):21-44.
  - 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 Christopher Zorn. 2001. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science* 45(October):951-67.
  - Cox, David Roxbee. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society, Series B* 34(2):187-220.

- Desmarais, Bruce A., and Jeffrey J. Harden. 2012. "Comparing Partial Likelihood and Robust Estimation Methods for the Cox Regression Model." *Political Analysis* 20(1):113-135. DOI:10.1093/pan/mpr042
- Grambsch, Patricia M., and Terry M. Therneau. 1994. "Proportional Hazards Tests and Diagnostics Based on Weighted Residuals." *Biometrika* 81(3):515-26.
- Grambsch, Patricia M., Terry M. Therneau, and Thomas R. Fleming. 1995. "Diagnostic Plots to Reveal Functional Form of Covariates in Multiplicative Intensity Models." *Biometrics* 51(December):1469-82.
- Keele, Luke J. 2010. "Nonproportionally Difficult: Testing for Nonproportional Hazards In Cox Models." *Political Analysis* 18:189-205.
- Licht, Amanda A. 2011. "Change Comes with Time: Substantive Interpretation of Nonproportional Hazards in Event History Analysis." *Political Analysis* 19(2):227-243.
- 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.
- 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* 18(Summer):155-95.

Day Five:

- **Morning: Survival Model Extensions**
  - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 9 & 11.
  - Box-Steffensmeier, Janet M., Roman Ivanchenko, and Christopher Zorn. 2006. "Cure Models for Political Science Research." Working paper: Ohio State University.

  Additional resources:

  - Banerjee, Sudipto, Melanie M. Wall, and Bradley P. Carlin. 2003. "Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota." *Biostatistics* 4(1):123-42.
  - Bennett, D. Scott. 1997. "Testing Alternative Models of Alliance Duration, 1816-1984." *American Journal of Political Science* 41(July):846-78.
  - Box-Steffensmeier, Janet M., and Suzanna De Boef. 2005. "Repeated Events Survival Models: The Conditional Frailty Model." *Statistics in Medicine* 25(December):3518-33. DOI: 10.1002/sim.2434.
  - Box-Steffensmeier, Janet M., Suzanna L. De Boef and Kyle A. Joyce. 2007. "Event Dependence and Heterogeneity in Duration Models: The Conditional Frailty Model." *Political Analysis* 15(3):237-256.
  - Box-Steffensmeier, Janet M., Peter Radcliffe, and Brandon Bartels. 2005. "The Incidence and Timing of PAC Contributions to Incumbent U.S. House Members, 1993-94." *Legislative Studies Quarterly* 30(November):549-79.

- Carpenter, Daniel. 2002. "Groups, the Media, Agency Waiting Costs and FDA Drug Approval." *American Journal of Political Science* 46(July):490-505.
- Chiozza, Giacomo, and Hein E. Goemans. 2004. "International Conflict and the Tenure of Leaders: Is War Still Ex Post Inefficient?" *American Journal of Political Science* 48(July):604-18.
- Hettinger, Virginia, and Christopher Zorn. 2005. "Explaining the Incidence and Timing of Congressional Responses to the U.S. Supreme Court." *Legislative Studies Quarterly* 30(February):5-28.
- Omori, Yasuhiro and Richard A. Johnson. 1993. "The Influence of Random Effects on the Unconditional Hazard Rate and Survival Functions." *Biometrika* 80(4):910-14.
- Maller, R. A. and S. Zhou. 1996. *Survival Analysis with Long-Term Survivors*. New York: Wiley.
- Manton, Kenneth G., Eric Stallard and James W. Vaupel. 1981. "Methods for Comparing the Mortality Experience of Heterogeneous Populations." *Demography* 18(August):389-410.
- Sastry, Naryan. 1997. "A Nested Frailty Model for Survival Data, With an Application to the Study of Child Survival in Northeast Brazil." *Journal of the American Statistical Association* 92(438):426-35.
- Schmidt, Peter and Anne D. Witte. 1989. "Predicting Recidivism Using 'Split-Population' Survival Time Models." *Journal of Econometrics* 40(1):141-59.
- Tsodikov, Alexander. 1998. "A Proportional Hazards Model Taking Account of Long Term Survivors." *Biometrics* 54:1508 15.
- Vaupel, James W., Kenneth G. Manton, and Eric Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography* 16:439-54.

- **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 github 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 an in-class final examination given in the afternoon of the last day (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.

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.

Workload

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