Advanced Maximum Likelihood: Survival Analysis

Summer 2015

Professor Christopher Zorn
Department of Political Science
Pennsylvania State University
E-mail: zorn@psu.edu
August 3-14, 2015

Description

The second two weeks of the "Advanced Maximum Likelihood" course is designed as an in-depth introduction to survival analysis: models for data where the response variable takes the form of a time to an event. Models for survival data are also known as "duration models" and "event history models," and have seen increasingly widespread use in the social sciences over the past thirty years. Modeling survival data presents a unique set of challenges.

The materials in this part of the course will concentrate on practical applications of models for survival data, including issues relating to data setup, software, and interpretation of model estimates. All homework exercises can be conducted using either Stata 14.0 or R 3.2.0 or thereabouts. All notes, exercises, and data will be available at the course github repo.

This syllabus is designed to provide an overview to the course. Clickable links are printed in Penn State blue.

Course Readings

Required Text/Materials

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. New York: Cambridge University Press.

Additional readings as necessary, all of which will be available on github and/or through JSTOR.

A Few Other Potentially Useful/Recommended Readings

Beck, Nathaniel. 1998. "Modeling Space and Time: The Event History Approach." In E. Scarbrough and E. Tanenbaum, eds., *Research Strategies in the Social Sciences*. London: Oxford University Press.

Box-Steffensmeier, Janet M. and Bradford Jones. 1997. "Time is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41(October):1414-61.

Broström, Göran. 2012. *Event History Analysis with R*. New York: Chapman & Hall.

Chung, Ching-Fan, Peter Schmidt, and Ann D. Witte. 1991. "Survival Analysis: A Survey." *Journal of Quantitative Criminology* 7(March):59-98.

Collett, Dave. 2003. *Modeling Survival Data in Medical Research*, 2nd Ed. London: Chapman & Hall.

Cox, David Roxbee and Dallin Oakes. 1984. *Analysis of Survival Data*. London: CRC/Chapman & Hall.

Fleming, Thomas R., and David P. Harrington. 2005. *Counting Processes and Survival Analysis*, Second Edition. New York: Wiley.

Hosmer, D., Stanley Lemeshow, and Susanne May. 2008. *Applied Survival Analysis: Regression Modeling of Time to Event Data*, Second Edition. New York: Wiley.

Ibrahim, Joseph G., Ming-Hui Chen, and Debajyoti Sinha. 2005. *Bayesian Survival Analysis*. New York: Springer.

Kalbfleisch, J. D. and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*, Second Edition. New York: Wiley.

Kiefer, Nicholas M. 1988. "Economic Duration Data and Hazard Functions." *Journal of Economic Literature* 26(June):646-79.

Kleinbaum, David G. and Mitchel Klein. 2011. *Survival Analysis: A Self-Learning Text*, Third Edition. New York: Springer-Verlag.

Klein, John P., and Melvin L. Moeschberger. 2005. *Survival Analysis: Techniques for Censored and Truncated Data*, Second Edition. New York: Springer-Verlag.

Lancaster, Tony. 1992. *The Econometric Analysis of Transition Data*. New York: Cambridge University Press.

Lawless, J. F. 2002. *Statistical Methods and Models for Lifetime Data*, Second Edition. New York: Wiley.

Some Other Useful Resources

The **Political Methodology Section** of the American Political Science Association was created to provide APSA members with an interest in political methodology with a forum in which to meet and discuss ideas. The section publishes a quarterly newsletter (*The Political Methodologist*), a quarterly journal on political methodology (*Political Analysis*), conducts a discussion list on topics relating to political methodology, and maintains an extensive electronic archive of papers, accessible via their homepage.

Also, the **Inter-University Consortium for Political and Social Research** (ICPSR), at the University of Michigan, maintains an extensive archive of data in the social and behavioral sciences. Much of it is acces-

sible via their homepage.

Software, Statistical and Otherwise

You are welcome to make use of whatever statistical software you choose to complete the homework exercises, so long as the manner by which your results are generated and conclusions reached are transparent. However, due to the limits of instructor time and patience, I will support only two software packages. Both are available on the machines in the political science computing labs.

R

R is a statistical environment and high-level programming language for data analysis and display. It is effectively the GNU version of the S language; as such, it is free (both as in speech and as in beer) and open source. The current (early-May 2015) version of R is 3.2.0. R is an *object-oriented* language; unlike Stata (and most other statistical packages), it operates by assigning values to objects in the workspace. In the notes, handouts, etc., R commands will be preceded by a caret (">"):

```
> my.results<-lm(Y~X)
```

The Comprehensive R Archive Network (CRAN) is the go-to spot for all things R-related. I cannot begin to list all the R-related resources available on the web; for newbies, however, it might be useful to check out the Introduction to R, this page in getting data into R, and the various R "cheat sheets" here, here, and here. Stata users who are interested in learning R should check out the Moving from Stata to R page at the R Project's wiki.

Course examples will almost always be developed and illustrated using R .

Stata

At the present time, Stata is probably the most widely-used statistical package in the social sciences. It is a powerful tool for data management, analysis, and display, and boasts some of the best manuals and on-line help of any existing software package. Stata is commercial software; the current version of Stata is 14.0, but previous versions (back to v. 9, at least) can also be used for the class. In the class notes, handouts, etc., Stata commands will appear in a fixed-width font and will be preceded by a period (".."):

```
. regress Y X
```

Stata newbies may want to check out:

Acock, Alan C. 2014. *A Gentle Introduction to Stata*, 4th Ed.. College Station, TX: Stata Press.

Beyond this, the Stata homepage is a valuable resource for questions about the Stata statistical software. There are a number of useful Stata references on the web, including Scott Long's page at IU and an excellent Stata "help page" sponsored by UCLA.

Course Schedule

Readings should be completed prior to coming to class on the assigned day. Note that we will not, in general, hew closely (or at all) to the readings themselves, other than topically. Links are generally to DOIs or to stable PDFs at JSTOR.

Day One: Introduction to Survival Models

- **Readings**
  - *Required*:
    - Hosmer, David W., and Stanley Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*, pp. 27-84 and Appendix 1.
    - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 1-2.
  - *Recommended*:
    - Cioffi-Revilla, Claudio. 1984. "The Political Reliability of Italian Governments: An Exponential Survival Model." *American Political Science Review* 78(2):318-37.
    - Zelditch, Morris Jr. and Joan Butler Ford. 1994. "Uncertainty, Potential Power, and Non-decisions." *Social Psychology Quarterly* 57(1):64-73.

Day Two: Parametric Survival Models

- **Readings**
  - *Required*:
    - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapter 3.
    - Alt, James, and Gary King. 1994. "Transfers of Governmental Power: The Meaning of Time Dependence." *Comparative Political Studies* 27(2):190-210.
  - *Recommended*:
    - 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.
    - McCarty, Nolan and Rose Razaghian. 1999. "Advice and Consent: Senate Responses to Executive Branch Nominations." *American Journal of Political Science* 43(October):1122-43.
    - Teachman, Jay D., and Mark D. Hayward. 1993. "Interpreting Hazard Rate Models." *Sociological Methods and Research* 21(February):340-71.

Day Three: Cox's Proportional Hazards Model

- **Readings**
  - *Required*:
    - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 4-5.
    - 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
  - *Recommended*:
    - Cox, David Roxbee. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society, Series B* 34(2):187-220.
    - Hegre, Havard, Tanja Ellingsen, Scott Gates, and Nils Petter Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816-1992." *American Political Science Review* 95(March):33-48.
    - Idris, Muhammad, and Christopher Zorn. 2015. "Proportional Hazards Analysis of Survival Data with Tied Survival Times: Theory and Best Practices." Working paper: Pennsylvania State University.
    - Leung, M. K., D. Rigby, and T. Young. 2003. "Entry of Foreign Banks in the People's Republic of China: A Survival Analysis." *Applied Economics* 35(1):21-31.
    - Lindsey, J. K. 1998. "Counts and Times to Events." *Statistics in Medicine* 17:1745-51.
    - Pevehouse, Jon. 2002. "With a Little Help from My Friends? Regional Organizations and the Consolidation of Democracy." *American Journal of Political Science* 46(July):611-26.

Day Four: Survival Analysis: Discrete-Time Alternatives

- **Readings**
  - *Required*:
    - 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).
    - 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.
  - *Recommended*:
    - 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.
    - Singer, Judith D., and John B. Willett. 1993. "Its 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: Survival Analysis: Diagnostics, etc.

- **Readings**
  - *Required*:
    - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapter 8.
    - 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.
    - Keele, Luke J. 2010. "Nonproportionally Difficult: Testing for Nonproportional Hazards In Cox Models." *Political Analysis* 18:189-205.
    - Zorn, Christopher. 2000. "Modeling Duration Dependence." *Political Analysis* 8(Autumn): 367-380.
  - *Recommended*:
    - Dunkler, D., M. Schemper, and Georg Heinze. 2010. "Gene Selection in Microarray Survival Studies Under Possibly Non-Proportional Hazards." *Bioinformatics* 26:784-790.
    - 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.
    - Heckman, James J. 1991. "Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity." *American Economic Review* 81(May):75-79.
    - Licht, Amanda A. 2011. "Change Comes with Time: Substantive Interpretation of Nonproportional Hazards in Event History Analysis." *Political Analysis* 19(2):227-243.
    - Ng'andu, N. H. 1997. "An Empirical Comparison of Statistical Tests for Assessing the Proportional Hazards Assumption of Cox's Model." *Statistics in Medicine* 16:611-26.
    - Warwick, Paul. 1992. "Rising Hazards: An Underlying Dynamic of Parliamentary Government." *American Journal of Political Science* 36(November):857-76.

Day Six: Extensions (Multiple Events and Competing Risks)

- **Readings**
  - *Required*:
    - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. Chapter 10 and pp. 148-154.
    - Box-Steffensmeier, Janet M., and Christopher Zorn. 2002. "Duration Models for Repeated Events." *Journal of Politics* 46(November):1069-94.
    - Janet M. Box-Steffensmeier, Janet M., Suzanna Linn, and Corwin D. Smidt. 2014. "Analyzing the Robustness of Semi-Parametric Duration Models for the Study of Repeated Events." *Political Analysis* 22:183-204.
    - Pintilie, Melania. 2007. "Analysing and Interpreting Competing Risk Data." *Statistics in Medicine* 26:1360-67.

- *Recommended*:
  - Cleves, Mario. 1999. "Analysis of Multiple Failure-Time Data with Stata." *Stata Technical Bulletin* 49:30-39.
  - Crowder, Martin. 2012. *Multivariate Survival Analysis and Competing Risks*. New York: Chapman & Hall/CRC.
  - David, H. A., and M. L. Moeschberger. 1978. *The Theory of Competing Risks*. New York: MacMillan.
  - Diermeier, Daniel, and Randy T. Stevenson. 1999. "Cabinet Survival and Competing Risks." *American Journal of Political Science* 43(4) October: 1051-68.
  - Gordon, Sanford C. 2002. "Stochastic Dependence in Competing Risks." *American Journal of Political Science* 46(January):200-17.
  - Kelly, Patrick J. and Lynette L-Y. Lim. 2000. "Survival Analysis for Recurrent Event Data." *Statistics in Medicine* 19:12-33.
  - Wei, L. J. and David V. Glidden. 1997. "An Overview of Statistical Methods for Multiple Failure Time Data in Clinical Trials." *Statistics in Medicine* 16:833-39.
  - Wolbers, Marcel, et al. 2014. Competing Risks Analyses: Objectives and Approaches. *European Heart Journal*.
  - Zorn, Christopher and Steven R. Van Winkle. 2000. "A Competing Risks Model of U.S. Supreme Court Vacancies, 1789-1992. *Political Behavior* 22(June):145-66.

Day Seven: Identification and Separation

- **Readings**
  - *Required*:
    - Heinze, G. and M. Schemper 2001. "A Solution to the Problem of Monotone Likelihood in Cox Regression." *Biometrics* 57:114-119.
    - Zorn, Christopher. 2005. "A Solution to Separation in Binary Response Models." *Political Analysis* 13(2):157-70.
  - *Recommended*:
    - Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1): 27-38.
    - Heinze, Georg, and M. Ploner. 2002. "SAS And SPLUS Programs To Perform Cox Regression Without Convergence Problems." *Computer Methods and Programs in Biomedicine* 67:217-223.
    - Heinze, Georg, and R. Puhr. 2010. "Bias-Reduced and Separation-Proof Conditional Logistic Regression with Small or Sparse Data Sets." *Statistics in Medicine* 29:770-777.
    - Kosmidis, Ioannis, and David Firth. 2009. "Bias Reduction in Exponential Family Nonlinear Models." *Biometrika* 96:793-804.
    - Kosmidis, Ioannism and David Firth. 2011. "Multinomial Logit Bias Reduction via the Poisson Log-Linear Model." *Biometrika* 98:755-759.

Day Eight: Extensions: Cure Models

- **Readings**
  - *Required*:
    - Box-Steffensmeier, Janet M., Roman Ivanchenko, and Christopher Zorn. 2006. "Cure Models for Political Science Research." Working paper: Ohio State University.
  - *Recommended*:
    - 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.
    - 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.
    - Maller, R. A. and S. Zhou. 1996. *Survival Analysis with Long-Term Survivors*. New York: Wiley.
    - Schmidt, Peter and Anne D. Witte. 1989. "Predicting Recidivism Using 'Split-Population' Survival Time Models." *Journal of Econometrics* 40(1):141-59.
    - Tsodikov, A. 1998. "A Proportional Hazards Model Taking Account of Long Term Survivors." *Biometrics* 54:1508 15.

Day Nine: Frailty Models

- **Readings**
  - *Required*:
    - 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.
    - Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 9 & 11.
    - 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.
  - *Recommended*:
    - 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.
    - 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.
- 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.
- 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.

Day Ten: Catch-Up, Wrap-Up, and Review

- **No readings.**