

Assignment 1: Event History Analysis
First draft due October 27, 2009

For this assignment, you will utilize data from the National Longitudinal Survey of Youth. The dataset (zipped) is located at <http://www.sarkisian.net/sc706/nlsysubset.zip> You can also use your own data, but please consult with me to make sure that your dataset would be appropriate.

- 1) Devise a theoretical argument you can test using event history analysis of these data. These data contain information on employment, marriage, and fertility. For this assignment, focus on non-repeatable, single-type events. I especially encourage you to deal with the fertility data although that will be a bit of a challenge: you will have to transform the birth dates information into something useful for you, either independent variables, or event variables – that should typically be done while the data are in still in wide format. Make sure to include both time-varying and time-invariant variables.
- 2) Start a running log that will contain the commands and output for all of your models and diagnostics, with brief comments. (Make sure you open a log file each time you open Stata!)

Discrete-time analyses:

- 3) Do the data management to prepare the data for discrete-time event history analysis.
- 4) Investigate the univariate distributions for all variables and the bivariate relationships between the dependent and each of the independent variables; explore transformations if necessary.
- 5) Estimate a discrete-time event history model, examine the shape of the baseline hazard for the model. Consider and evaluate alternative ways to represent time in your model.
- 6) Perform the diagnostics – test whether the effects are time-invariant, check linearity, additivity, and multicollinearity, and outliers and influential observations. If necessary, consider remedies.
- 7) Evaluate the overall fit of your final model. Consider making your model more parsimonious by omitting some variables; evaluate how that affects the fit of your model.
- 8) Interpret the final results using exponentiated coefficients as well as predicted probabilities.

Continuous-time analyses:

- 9) Do the data management to prepare the dataset for continuous-time analysis and use stset.
- 10) Use nonparametric techniques to examine your data.
- 11) Estimate Cox proportional hazards model, examine the baseline functions for the model.
- 12) Consider whether it would be appropriate to use a parametric model, and if so, determine the type of the model to use. If you decided that a parametric model is appropriate, estimate it.
- 13) For your final model (either Cox or a parametric model), perform the diagnostics – test for proportionality, linearity, additivity, and outliers and influential observations. If necessary, consider remedies.
- 14) Evaluate the overall fit of your final model. Consider making your model more parsimonious by omitting some variables; make sure to evaluate how that affects the fit of your model.
- 15) Interpret your final model using exponentiated coefficients as well as survival functions for some meaningful sets of values. Write up your interpretation of the results of your final model. Compare the results with those you obtained using discrete-time event history analysis.
- 16) There is no page limit for your assignment but please edit it to contain only the relevant commands and output, and include the relevant graphs as well (you can copy and paste them into your word processor). Please submit your two dataset files (one for discrete-time analysis, one for continuous-time analysis) along with your assignment.

Journal write-up component (optional):

Write up the results like you would for a journal publication. First, include an Introduction that will provide a short substantive description of your theoretical argument, your research questions and hypotheses (1 page max.). Second, include a brief Data and Methods section (1-2 pages) describing the variables and the analysis methodology. Include any discussion of diagnostics and modifications in this section, either in the text or in the footnotes. Also, include a table with summary statistics for the variables you use (means, standard deviations, number of observations). Third, provide a 3-4 page description of the results including tables (in journal format) and any graphs assisting in the interpretation of results (do not show all tables and graphs – select those most helpful for understanding the results; it is typically not a good idea to show any diagnostic graphs). If it is absolutely necessary that you discuss some diagnostics or model modifications here rather than in the methods section, do so using footnotes. Finally, include a brief conclusion summarizing your findings and discussing contributions and limitations of your research (1 page max.). The page limit for this write-up is 8 pages double-spaced.

Assignment 1 Grading Sheet
Total Preliminary Grade: out of 100

- A. Theoretical argument (5 points):
- B. Discrete-time model (40 points):
 - 1. Data management to prepare the data for discrete-time analysis (5 points):
 - 2. Univariate and bivariate data examination (5 points):
 - 3. Discrete-time model estimation, baseline hazard, and parametrization of time (5 points):
 - 4. Test of time invariance (5 points):
 - 5. Testing multicollinearity, linearity, additivity, outliers, and selecting remedies (5 points):
 - 6. Evaluation of model fit and parsimony (5 points):
 - 7. Interpretation of the coefficients (5 points):
 - 8. Interpretation of predicted probabilities and the survival functions (5 points):
- C. Continuous-time model (50 points):
 - 1. Data management to prepare the data for continuous-time analysis (5 points):
 - 2. Using nonparametric techniques to examine the data (5 points):
 - 3. Cox model estimation and baseline curves (5 points):
 - 4. Considering a parametric model (5 points):
 - 5. Evaluation of model fit and parsimony (5 points):
 - 6. Test of proportionality assumption (5 points):
 - 7. Testing linearity, additivity, outliers, error terms, and selecting remedies (5 points):
 - 8. Interpretation of the coefficients (5 points):
 - 9. Interpretation of the survival functions (5 points):
 - 10. Comparison to the discrete-time results (5 points):
- D. Log organization (5 points):

Journal-style Write-up Grading Sheet
Total Preliminary Grade: out of 100

- 1. Introduction (15 points)
- 2. Data and methods (30 points)
- 3. Tables and graphs (15 points)
- 4. Description of results (30 points)
- 5. Conclusion (10 points)