UNIVERSITY OF ESSEX
INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH
Professor Stephen P. Jenkins <stephenj@essex.ac.uk>

Essex Summer School course 'Survival Analysis'
and
EC968. Part II: Introduction to the analysis of spell duration data

Lesson 9. Assorted other topics

Congratulations – you've made it this far! ☺

This document is currently a place holder, because I don't have any additional lesson material at present!

The Lessons included so far reflect the coverage of my Lecture course. If the scope were to be extended, then several potential topics suggest themselves. They might include the following:

- More about statistical inference and the properties of different estimators;
- Specification and goodness of fit tests;
- Examples and applications using data generated from more complicated sampling schemes (e.g. left truncation = delayed entry = stock sampling with follow-up).
- Multiple or repeated spell models. (As long as one assumes full independence of spells, then this is a relatively trivial topic – just pool the spells. But if one allows for correlated unobservables, then things quickly get rather harder …)
- More about construction and organisation of data sets using rather messier examples than in the Lessons so far e.g. more manipulation of dates; more manipulation of event histories; more use of real-life data from e.g. BHPS etc.

Other ideas include

- a case study worked through in some detail.
  In the Lectures we compare and contrast selected pairs of published articles. For practical applications using Stata, things are a little harder. Nonetheless the dropout data from Yamaguchi (1990) have not been utilised at all so far in the Lessons. One could perhaps use them to replicate Yamaguchi's examples, or to practice a range of other techniques.

Good luck. And please email me your comments and suggestions.

Stephen P. Jenkins <stephenj@essex.ac.uk>

PS Turn over for some notes on "Which model should I choose?"

Which model should I choose?

This is one of the most common questions that I get in class and, sadly, one for which there is no single definitive answer!

Here is a brief list of issues to take into account and suggestions. Read them in conjunction with Section 2.3 in Chapter 2 of the Lecture Notes ("Choosing a specification for the hazard rate").

- Are your survival time data continuous, grouped (interval-censored), or intrinsically discrete? (Inevitably, virtually all social science data are grouped, but whether this matters – the choice between a continuous or interval-censored specification – partly depends, broadly speaking, on the width of each interval relative to the average spell length. If spell lengths are recorded in months, is the typical spell length less than one year (in which case grouping may well an important issue) or about a decade (in which case grouping is likely to be less relevant).
- Do behavioural models suggest particular hypotheses about the shape of the hazard in your case?
- Explore the shape of the hazard and related functions using non-parametric estimators (Lesson 3) – do these provide clues about whether e.g. the hazard is monotonically increasing or decreasing or what? Note too the informal graphical checks for the proportional hazards and log-logistic models.
- (Related) The shape of the hazard may also be explored in a regression framework using piecewise constant exponential models or their discrete time counterparts, using judiciously chosen partitions of the survival time axis. (Of course, if you are not interested in estimating how the hazard varies with survival time, you could simply use the Cox model for continuous time data, or the corresponding non-parametric model for grouped/discrete data.)
- (Related) Use a general-to-specific modeling strategy: start with a parametric model that characterizes the hazard in a relatively flexible way (e.g. generalized Gamma model for continuous time data), and then test whether the specifications nested within this are appropriate using likelihood ratio tests. (For non-nested models, alternative testing methods are required, based on e.g. AIC or BIC criteria.)
- Assess the goodness of fit of your model and check for outlier observations using residual analysis. This topic is not currently covered in the Lessons (reflecting the practice of most applied economists!) See the discussion in the Stata Reference Manuals. If you are using a proportional hazards model applied to continuous time data, see also the procedures for checking the appropriateness of the proportional hazards assumption in a regression context.
- In addition to these checks of regression models, there are also the usual specification checks that you can undertake regarding the exclusion or inclusion of various covariates, based on Wald or likelihood ratio tests.