Advanced Topics in Maximum Likelihood Estimation:
Duration Analysis/Event History Analysis
Professor Brad Jones
University of California, Davis
Weeks 1-2

Contact Information: Professor Brad Jones
Department of Political Science
University of California, Davis
Davis, CA 95616
USA
E-mail: bsjjones@ucdavis.edu
Website: http://psfaculty.ucdavis.edu/bsjjones

Course Description
This section of the advanced MLE course will cover methods and models for duration data. Duration data record the length of time until some event occurs, for example, the termination of a cabinet government or the time until an unemployment spell ends. Because time-to-event occurrence is an important feature of these kinds of data, methods suitable to duration data are often referred to as event history analysis. In this course, we will consider a wide variety of event history modeling methods. Students will be asked to complete some problem sets that will involve estimating and interpreting event history models. In addition to consideration of duration models, we will take a “side-trip” and consider some non-traditional (i.e. not widely used) categorical models and consider their applicability to duration data. “Tutorials” will be available to students that describe some implementation issues pertinent to these models. Additionally, some lecture notes will be available as well. Any material necessary for downloading will be available at my website: http://psfaculty.ucdavis.edu/bsjjones. From here, you will be able to follow a link to this course to access tutorials, lecture notes, and article manuscripts.

Readings
The primary texts will be Box-Steffensmeier and Jones’ *Event History Modeling: A Guide for Social Scientists* (Cambridge University Press, 2004) and Kleinbaum and Klein’s *Survival Analysis: A Self-Learning Text* (Spring, 2005). In the daily itinerary of topics, several articles will be assigned (and will be posted on the website for the book). You will need to consult the website in conjunction with the date of the class to retrieve the readings.

Requirements
Students are expected to do the assigned readings and pay attention in class. There will be two or three short problem sets. The problem sets will entail estimation and interpretation of a variety of duration models. We will make use of both the R computing environment and Stata. Lecture notes and tutorials will be available on my website. Here, relevant code for both R and Stata can be found. Finally, students will be asked to turn in a short (1-2 page) research prospectus that outlines a research question(s) and hypotheses that could be appropriately tested using duration modeling techniques.
Itinerary
My principal goal is to give you an introduction to the fundamental elements of duration modeling and then consider in some detail parametric, non-parametric (via the Cox model), and “discretized” duration models for single-event and multi-event duration data. I do not assume any prior knowledge of event history modeling, though I obviously will assume knowledge of the basic principals of maximum likelihood estimation as well as a thorough understanding of the classical linear model and traditional binary link models (like logit or probit).

The following gives you the day-by-day itinerary of topics. There are two “classes” of readings each day: core and application. It is important that the core readings be completed in their entirety. Several applications are listed for each day's topics. You will not have time to read each application; I recommend choosing a couple that may be of interest to you. Just about all of the application readings (as well as the core articles) are available from J-Stor (http://www.jstor.org). Applications are highly useful to read and I encourage you to read as many of these as you can. The readings below are primarily drawn from the social sciences.

---

DAY 1: Monday, July 22
Preliminaries: Event History Data and the “Moving Parts” of Event History Analysis and an Introduction to Modeling Strategies

Core Readings:
• Box-Steffensmeier and Jones, Chapters 1—2.
• Hosmer et al, Chapters 1-3.

DAY 2: Tuesday, July 23
Describing and Modeling Duration Data: The Kaplan-Meier Estimator (and related estimators) and Parametric Models

Core Readings:
• Box-Steffensmeier and Jones, Chapter 3.
• K and K: Chapters 1 and 2

DAY 3: Wednesday, July 24
Estimation and Model Selection Issues in the Application of Parametric Duration Models

Core Readings:
• Box-Steffensmeier and Jones, Chapter 3.
• K and K, Chapter 7

DAY 4: Thursday, July 25
The Cox Proportional Hazards Model
Core Readings:
• Box-Steffensmeier and Jones, Chapter 4.
• K and K, Chapter 3

DAY 5: Friday, July 26.
The Proportional Hazards Property and other Cox Model Diagnostics and Other Issues

Core Readings:
• Box-Steffensmeier and Jones, Chapter 8.
• K and K, Chapters 3-5.

DAY 6: Monday, July 29
“Discretized” Duration Data and Associated Models

Core Readings:
• Box-Steffensmeier and Jones, Chapter 5.

DAY 7: Tuesday, July 30
Models for Competing Risks: Discrete and Cox

Core Readings:
• Box-Steffensmeier and Jones, Chapter 10.
• K and K, Chapter 9

DAY 8: Wednesday, July 31
Repeatable Events

Core Readings:
• Box-Steffensmeier and Jones, Chapter 10.
• K and K, Chapter 9
DAY 9: Thursday, Aug. 1  
Frailty and Split-Population Models  

Core Readings:  
• Box-Steffensmeier and Jones, Chapters 9.  

DAY 10: Friday, Aug. 2  
Special Topics in Duration Analysis  

Readings: TBA
Course Description

These two weeks of the advanced MLE course will focus on methods and models for panel and time-series cross-section (TSCS) data. These types of data occur when we have observations for multiple units collected at multiple points in time. Panel data typically are sample data in which there is a large number of cross-sectional units and few time points. TSCS data are typically data in which the units are of interest in themselves, the number of time points is much larger than in panel data, and the number of time points is either larger or approximately equal to the number of cross-sectional units. Since TSCS data are more widely used in political science outside of the area of panel survey studies, much of these weeks will focus on TSCS data rather than panel data.

This portion of the class will cover several questions that are central to the use of TSCS and panel data. Among these are fixed effects and random effects models, dynamic models, random coefficient models, models for limited dependent variables, models for spatial dependence, and panel attrition. Students will be asked to complete problem sets that involve estimating models for data collected in time and space. Because the focus of this class is on providing students with knowledge of methods they can apply in their own research, students will also be asked to write a short research proposal that describes a research question and hypotheses that can be tested using the methods learned in this course.

Readings

The primary text is Cheng Hsiao’s *Analysis of Panel Data*, 2nd Ed. (Cambridge University Press, 2003). There are also several required articles during this course. You should read both the required pages in Hsiao and the articles before the portions of the course for which they are assigned.

Requirements

Although not a formal prerequisite, familiarity with statistics at the level of a linear algebra treatment of regression is highly advisable for students enrolled in the course. In addition to the required reading, students will complete a few short problem sets during the course. We will make use of both Stata and R during the course. In addition to these problem sets, students will write a short (1-2 page) research proposal that describes a research question, testable hypotheses,
and data that might be used to test these hypotheses for a research question of interest to the student.

**Course Outline**

The schedule below lists the topics that will be covered during these two weeks. To provide flexibility in the coverage of this material, I don’t provide specific dates for this material, but we will follow the order below in covering these topics.

1. **Course Introduction/Panel vs. TSCS Data**

   Required Reading:

   Hsiao, Chapter 1

2. **Heterogeneity and Pooling/Analysis of Covariance**

   Required:

   Hsiao, Chapter 2


3. **Fixed Effects Models**

   Required:

   Hsiao, Chapter 3


   Recommended:


4. **Random Effects Models**

   Required:

   Hsiao, Chapter 3

   Recommended:

5. Panel-Corrected Standard Errors

Required:


Recommended:


6. Variable Coefficient Models/Rarely Changing Variables

Required:

Hsiao, Chapter 6


Recommended:


7. Modeling Dynamics in TSCS Data

Required:


Recommended:


**8. Models for Categorical and Limited Dependent Variables/Panel Attrition**

Required:


Hsiao, Chapters 7 and 8.

Recommended:


**9. Spatial Models for TSCS Data**

Required:

