Survival Models

A. Introduction

1. Economists are interested in a variety of problems that require them to examine the determinants of the length of time, or the duration, that an individual spends in a particular state; examples of problems include
   a. Spells of unemployment or employment
   b. Spells of welfare receipt
   c. Spells of business operation
   d. Birth timing and spacing
2. Definitions of spells and events
   a. Assume that there are two (or possibly more) mutually exclusive states or conditions that a person could occupy (for instance, the person could be employed or unemployed)
   b. A *spell* refers to the time that a person spends in one state or condition before transitioning to another
   c. All spells are characterized by start and ending times; we refer to the event of transitioning from the initial state to the other state—the event of leaving or ending a spell—as a *spell exit*
   d. We refer to the amount of time that the person spends in the spell as the *duration*
   e. There are several equivalent ways to examine spell behavior; specifically, we can examine the duration of spells, the timing of exits, or other characteristics

3. Duration distribution

   a. Assume that the spell duration for a person is a random variable, T, which has a cumulative distribution $F(t) = \text{Prob}(T \leq t)$

      1) If T is a continuous random variable, then it also has an associated density function, $f(t)$

      2) If T is a discrete random variable, then it has an associated probability function, $p(t) = \text{Prob}(T = t)$

   b. We are interested in describing the distribution of T

4. Simple models

   a. Assume that we have data on completed spells for a sample of people; by complete, we mean that we know the start and ending times and hence the *T*s for all the spells

   b. Suppose also that we have information on other (time-invariant) characteristics of the people, $\boldsymbol{X}$

   c. We could examine the association of these characteristics with average spell lengths by estimating a regression of the form

$$\text{E}(T \mid \boldsymbol{X}) = \alpha + \boldsymbol{\beta}'\boldsymbol{X}$$

   d. We could also estimate other models to examine other parts of the distribution

      1) Median (least absolute deviations) regressions to examine conditional medians, or more generally quantile regressions

      2) Discrete choice models to examine the conditional probabilities of spells lasting past given points

5. Some data issues

a. Censoring

1) Information about some spells may be incomplete
2) When we collect data, some spells may already be in progress, and we might not know the start date; we refer to these as *left-censored spells*
3) Also, some spells might not end by the end of our observation window; also, people could drop out of our sample before an exit is observed; we refer to these as *right-censored spells*
4) Spells can also be doubly censored
5) In each case, we know that the spell is longer than the duration that we observe
6) We cannot treat these observations as completed spells; we know that the true spell durations are longer, and treating the censored observations as if they were complete will systematically understate the actual distributions
7) Dropping the observations is also problematic
   a) Long spells are more likely to be censored than short spells
   b) Dropping censored spells produces a sample that contains a disproportionate number of short spells
   c) Again, spell distributions would tend to be understated

8) Event history procedures typically address censoring problems

    a) Most address right-censoring

    b) Procedures also available to address left-censoring; these procedures, however, are more complicated and less frequently used (for an example see, Moffitt and Rendell 1995); usual practice is to either

        i) Drop ongoing (left-censored) spells, or

        ii) Use very restrictive models such as simple Markov models or approximate corrections (see, e.g., Ribar 2005)

b. Time-varying covariates

    1) Observed characteristics for people may change over the course of their spells—for example, local employment conditions could improve, their health might worsen, etc.

    2) Simple models cannot incorporate time-varying covariates but most event-history models can

c. Duration dependence

    1) We often want to examine whether people become more or less likely to exit a spell the longer they stay in a spell

    2) If the probability of exit changes with the length of a spell, the spell is said to exhibit *duration dependence*

    3) Consider standard job search models

a) Models in completely stationary environments predict that there will be no change in the probability of leaving unemployment over the course of a spell (no duration dependence)

b) Models with finite time horizons, time-limited unemployment insurance benefits, and borrowing constraints predict that people will become less choosy over time and more likely to leave unemployment (positive duration dependence)

c) Models with skills depreciation predict that people will become less capable of finding work as their unemployment spells progress (negative duration dependence)

4) It is difficult to characterize duration dependence using simple models

5) The problem is that duration dependence is a characteristic of the distribution function $F(t)$

a) Many standard procedures, such as regression models, only look at one point in the distribution; an analysis of duration dependence requires an examination of multiple points

b) Event history procedures typically use either convenient function forms for $F(t)$ or other restrictions that make it easier to examine duration dependence

B. Non-parametric Methods for Describing Discrete-Time Data

1. Assume that the durations of spells are measured in discrete intervals, such as hours, days, weeks, months, or years
   a. The random variable describing the spell length is discrete
      1) It takes on the values 1, 2, 3, … t_{Max}
      2) With associated probabilities p_1, p_2, p_3, … p_M,
      3) And with a CDF, $F(t) = p_1 + p_2 + \dots p_t$
   b. Let n_1 denote the number of individuals who leave the sample in the first period, n_2 denote the number of individuals who leave in the second period, and so on; let N denote the total number of individuals
   c. If there is no censoring (all people leave the sample because of transitions), the probabilities can be estimated from the proportions of individuals observed to leave at each period, e.g., $\hat{p}_j = n_j / N$, for j = 1, t_{Max}
2. Kaplan-Meier approach extends this to consider censoring
3. Some additional notation is helpful
   a. Let $S(t) = 1 - F(t) = \text{Prob}(T > t)$ denote the *survivor function*; this is simply the probability that an individual transitions after (survives past) time t; this is useful for summarizing what we know about censored observations
   b. Let $\lambda(t) = p(t) / S(t-1)$ be the *hazard function*; this is the probability that someone who survives up to t (survives past t–1) transitions at exactly t

c. Distinguish between people who leave in each period because they transition out, h_j, and people who leave because they are censored, m_j. Using our previous notation, $n_j = h_j + m_j$

4. Estimates of the survivor function can be constructed

$$\hat{S}(t-1) = \frac{1}{N}\sum_{j=t}^{t_{Max}} n_j$$

5. And estimates of the hazard function can be constructed

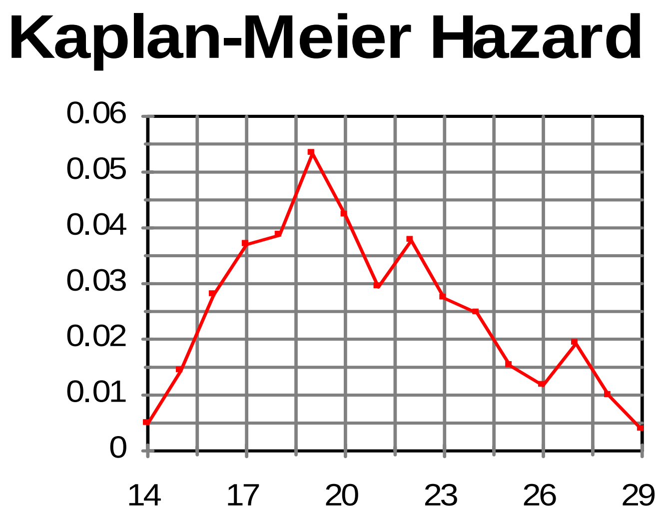
$$\hat{\lambda}(t) = h_j / \sum_{j=t}^{t_{Max}} n_j$$

6. Example, consider the following data on transitions to premarital births using data from the NLSY79

Age	Women at risk	Premarital births	KM hazard
14	2035	10	.0049
15	2017	29	.0144
16	1962	55	.0280
17	1864	69	.0370
18	1707	66	.0387
19	1500	80	.0533
20	1277	54	.0423
21	1088	32	.0294
22	927	35	.0378
23	767	21	.0274
24	647	16	.0247
25	524	8	.0153
26	426	5	.0117

27	363	7	.0193
28	303	3	.0099
29	259	1	.0039

In this example, censoring occurs because some women stop being respondents in the sample (attrition) and because some women marry before giving birth



7. Kaplan-Meier hazard functions are the primary way of conducting descriptive analyses of spell data

8. A survivor-based approach is available for characterizing continuous time data

C. Parametric distributions for continuous-time spells

1. The survival function has a similar definition in continuous time: $S(t) = \text{Prob}(T \geq t) = 1 - F(t)$

2. The definition of the hazard rate differs

$$
\begin{aligned}
\lambda(t) &= \lim_{\Delta t \to 0} \frac{\text{Prob}(t \le T < t + \Delta t \mid T \ge t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t \, S(t)} \\
&= \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}
\end{aligned}
$$

3. Survival functions and hazard functions are clearly related to underlying density and distribution functions
4. In working with these functions it is useful to note the following relationships
   a. Relationship between hazard and survival function

   $$\lambda(t) = \frac{-\,\mathrm{d} \ln S(t)}{\mathrm{d}t}$$

   b. Relationship between density and hazard function

   $$f(t) = S(t)\, \lambda(t)$$

   c. Integrated hazard function

   $$\Lambda(t) = \int_0^t \lambda(s)\, \mathrm{d}s$$

   d. Other relationships

   $$S(t) = e^{-\Lambda(t)} \quad \text{and} \quad \Lambda(t) = -\ln S(t)$$

5. Some common parametric distributions for continuous-time data

Distribution	Hazard/density function	Survival function
Exponential	$\lambda(t) = \lambda$	$S(t) = e^{-\lambda t}$
Weibull	$\lambda(t) = \lambda p(\lambda t)^{p-1}$	$\ln S(t) = -(\lambda t)^p$
Gompertz	$\ln \lambda(t) = \alpha + \beta t$	$\ln S(t) = -(\lambda(t) - e^{\alpha}) / \beta$
Log normal	$f(t) = (p/t)\, \varphi\, [p \ln(\lambda t)]$	$S(t) = \Phi[-p \ln(\lambda t)]$

6. Maximum likelihood estimation
   a. likelihood function is

$$\begin{aligned}\ln L(\theta) &= \sum_{\text{uncensored}} \ln f(t \mid \theta) + \sum_{\text{censored}} \ln\left[1 - F(t \mid \theta)\right] \\ &= \sum_{\text{uncensored}} \ln \lambda(t \mid \theta) + \sum_{\text{all}} \ln S(t \mid \theta)\end{aligned}$$

   b. essentially the same likelihood function as the censored regression model (similar data issues)

D. Heterogeneity bias
1. So far, we have examined parametric and non-parametric means of describing the time pattern of exits from spells
2. We have assumed that the spells are drawn from a single distribution with a common probability function – we have not allowed for heterogeneity in the hazards for individuals
3. Presence of heterogeneity leads to bias in the estimates of duration dependence
4. Consider the following example

a. Assume that there are two types of people: one type with exponential hazard λ_1 (leavers) and another with exponential hazard λ_2 (stayers), where $\lambda_2 < \lambda_1$

b. Next consider what happens to the average hazard as spells progress

c. As spells progress, the "leavers" will exit faster than the "stayers" and the sample of survivors will become disproportionately composed of "stayers"

d. With controls for heterogeneity, there will appear to be negative duration dependence

e. The failure to account for heterogeneity generally leads to bias (this is different from standard regression models where only certain types of heterogeneity lead to bias)

f. The intuition here is that event history models fit the distribution of observed exit times

   1) unless you specify otherwise, the models assume that all of the heterogeneity in exits is associated with duration dependence

   2) the models conflate different sources of heterogeneity

   3) similar to issues that arise in other ML models

E. Multivariate models – accounting for observed heterogeneity

   1. General approaches for parametric models – technically any parameter could be specified to be a function of observed variables (e.g., for the exponential model, could specify $\lambda = e^{\beta' X}$)

2. Proportional hazard specifications

   a. A more common way that observed heterogeneity is incorporated is through a "proportional hazards" assumption

      1) let $\lambda(t, X; \beta) = l(t)\, e^{\beta' X}$ where $l(t)$ is a *baseline hazard* and $e^{\beta' X}$ is a proportional shifter

      2) in this case, the observed characteristics shift the entire hazard function up or down

   b. This greatly simplifies the calculation of the hazard but is very restrictive

      1) Restricts all individuals to have hazards with the same shapes

      2) Assumption should be tested

3. Discrete logistic hazard

   a. Consider a hazard evaluated at discrete intervals

   b. Assume that the hazard follows a log-odds specification such that

$$\lambda(t) = \frac{\exp(\beta' X_t)}{1 + \exp(\beta' X_t)}$$

   c. Suppose that transition occurs at duration t; construct observations so that

      1) Observations for first $t-1$ periods have outcomes $Y_j = 0$ ($j = 1, t-1$)

      2) Observation for the t^{th} period has outcome $Y_t = 1$

3) Put another way, construct t observations for each person with a sequence of Y_j values equal to 0, 0, 0, 0, …, 0, 1

d. If the observation is right-censored at t, construct a sequence with all zeros

e. Each outcome would also be accompanied by measures of the observed characteristics in each time period X_j, for $j = 1, t$

f. With these data, you could then run a standard logit model in which the Y_j values are the dependent variables and the X_j are the independent variables

g. Specification is easy to implement and VERY flexible

1) Easy to specifiy flexible baseline hazard function

2) Incorporates time-varying covariates

3) Can be estimated with standard logit software once the data are transformed

4) Workhorse specification for exploratory data work

h. Besides the logistic distribution, discrete-time hazards are often fit with the complementary log-log distribution

1) For this distribution, the hazard function is

$$\lambda(t) = 1 - \exp[-\exp(\beta' X_t)]$$

2) For this specification, you would set up the data in the same way as for the discrete-time logit model

4. Cox partial likelihood model

a. One general weakness of parametric event history models is that they often restrict the shape of the hazard function; like other distributional misspecifications, this can lead to biased estimates

b. The Cox partial likelihood approach is a semi-parametric approach that addresses this concern

   1) Combines a non-parametric specification of the baseline hazard with

   2) A proportional hazards assumption

c. General description of approach when there are complete data

   1) Arrange the observations in the order that the transitions occur from t_1 to t_n

   2) At any point in time, t_j, the risk set for departures will be $R(t_j)$, the observations t_j through t_n

   3) Given the sequence up to t_j (t_1, t_2, ..., t_j), that is the ranking of transition times, the probability one transition occurs at t_j out of the risk set, $R(t_j)$, is

$$\frac{\lambda(t_j, X_j; \beta)}{\sum_{k \in R(t_j)} \lambda(t_j, X_k; \beta)} = \frac{\exp(\beta' X_j)}{\sum_{k \in R(t_j)} \exp(\beta' X_k)}$$

   This expression does not depend on the baseline hazard

d. The expression can be modified to account for right censoring and for discrete exit times (ties in the exit times)

e. Estimates of β are obtained conditional on an arbitrary baseline hazard function

f. Note: uses a partial, or conditional, likelihood approach, rather than a full likelihood approach

g. Even more flexible versions, such as models with stratified baseline hazard (separate baseline hazards for different groups), are available

5. Stata commands for descriptive and multivariate event-history analyses

a. Start by creating an event history data set

1) If there is just one record for each failure time (no time-varying covariates), use

stset *time_variable*, **failure**(*failure_variable*)

where *time_variable* gives the time to failure or transition and *failure_variable* is an indicator for whether a failure occurred (the alternative is that the observation is right censored)

2) When multiple records are available for each spell (e.g., if there are time-varying covariates), add an **id**(*id_variable*) option to the **stset** command and use the *time_variable* to indicate the end of each interval

3) Once this is done, Stata estimates models and calculates statistics in the context of this data arrangement (do not need to specify dependent variables)

b. Kaplan-Meier estimates for descriptive analyses

1) For survival functions, use the **sts** command:

sts OR **sts graph**

will create a graph with the KM survivor function; can produce list output by substituting **list** for **graph**; can perform conditional analyses by adding **by(***conditioning_variable***)** option

2) Can also produce smoothed hazard estimates by adding the **hazard** option:

sts, hazard

c. Estimating parametric models

1) Use **streg** command:

streg *list_of_variables*, **distribution**(*dist_type*)

2) Where the distributions include

exp	exponential
weibull	Weibull
gamma	generalized Gamma
gompertz	Gompertz
lnormal	log normal
llog	log-logistic

3) Unless you specify otherwise, Stata will output exponentiated coefficients (interpreted as hazard ratios); to get regular (untransformed) coefficients use **nohr** option

d. Estimating the Cox partial likelihood model

stcox *list_of_variables*

e. Estimating discrete-time logistic model

1) Can use **pgmhaz8** module written by S. Jenkins

2) Program estimates a discrete-time complementary log-log model
3) Use **ssc install pgmhaz8** to install program
4) Syntax is

pgmhaz8 *list_of_variables*

5) Reports results from standard complementary log-log model and from a model that accounts for gamma-distributed unobserved heterogeneity (frailty)

F. Discrete Outcomes and Interval Outcomes

1. We often have data that are reported in countable units such as days, weeks, or months; an issue arises, however, with whether the transitions actually occur on these types of time scales
   a. Discrete outcomes actually occur at discrete intervals; examples include some types of programmatic outcomes which terminate at the end of a day or month; here the underlying event history process is discrete
   b. Interval measures are used when the exact date of a transition is unknown but a range of dates is known; here the underlying transition is *continuous* but the measures are discrete
   c. We use different models for these two types of data
2. Discrete outcomes can be modeled using a discrete procedure, such as the logistic or conditional log-log procedures described earlier

3. Interval outcomes should be modeled using a modified continuous procedure
    a. Suppose that a transition was known to occur on or after duration t_j but before duration t_k ($t_j < t_k$)
    b. The hazard would be

$$\lambda(t_j \le T < t_k) = \frac{F(t_k) - F(t_j)}{S(t_j)}$$

    c. The aML software accommodates interval outcomes; stata does not appear to address this issue (treats ending times as exact exit times)
4. Time-varying explanatory variables
    a. If explanatory measures change values *at discrete points in time*, we can use a variant of the interval estimator to accommodate them
    b. Essentially, this requires breaking any spell into discrete subperiods corresponding to times when the explanatory variables remain constant
    c. Stata can accommodate this using the multiple-record version of the **stset** command
    d. This approach does not work if the explanatory measures are continually changing (i.e., are a continuous function of duration); for continually changing measures you would have to include the duration dependence in the specification of the hazard

G. Unobserved Heterogeneity

1. Suppose that the hazard rate depended on a set of observed variables, X, and an unobserved variable, μ, such that

$$\lambda(t \mid \mu) = \lambda(t, X \mid \beta, \mu)$$

2. In particular, assume that μ is a continuous random variable with a density function $g(\mu)$

3. If μ was observed, we could just include it as an explanatory variable; however, because μ is unobserved we must condition on all possible realizations of μ; this is equivalent to examining the expected value of $\lambda(t)$ or

$$E[\lambda(t)] = \int \lambda(t \mid \mu) g(\mu)\, \mathrm{d}\mu$$

4. For some specifications of $g(\mu)$ and $\lambda(t \mid \mu)$, it is possible to derive a closed-form expression for this expected value

   a. The gamma distribution is one such distribution; it is especially convenient to work with

   b. The Stata **streg** procedure will estimate models with this type of unobserved heterogeneity by adding the **frailty(gamma)** option

   c. Jenkins' **pgmhaz8** discrete-time complementary log-log procedure also estimates models with gamma-distributed unobserved heterogeneity

5. If μ is normally distributed, a closed-form expression is not possible; however, Gauss-Hermite quadrature can be used to give an accurate approximation

$$\int \lambda(t \mid \mu) g(\mu)\, \mathrm{d}\mu \approx \sum_{j=1}^{M} \lambda(t \mid \alpha_j) g(\alpha_j) w_j$$

where α_j is an abscissa, w_j is a weight, and M is the number of quadrature points (see earlier notes on numerical methods)

a. Logit and complementary log-log models with normally-distributed heterogeneity can be estimated in Stata by using the longitudinal/panel procedures for these models (**xtlogit** and **xtcloglog**) with the random effects (**re**) options

b. Logit and piecewise linear Gompertz models with normally distributed unobserved heterogeneity can also be estimated using the aML software package

6. An alternative is to assume that μ follows a discrete distribution with M outcomes $\mu_1, \mu_2, \ldots \mu_M$ and M probabilities $\pi_1, \pi_2, \ldots \pi_M$; then

$$\mathrm{E}[\lambda(\mathrm{t})] = \sum_{j=1}^{M} \lambda(t \mid \mu_j)\pi_j$$

a. We refer to this specification as a *finite mixture* model (Heckman & Singer 1984)

b. Some normalization is needed such as $\mu_1 = 0$ and $\pi_1 = 1 - \pi_2 - \ldots - \pi_M$

c. The specification is very flexible and can be viewed as an approximation to an arbitrary distribution

d. Note the similarity between this expression and the Gauss-Hermite quadrature expression

   1) the Gauss-Hermite quadrature specification is clearly a special case

2) comparison helps to show how the finite mixture approach generalizes the specification of the distribution

e. Stata software

1) S. Jenkins has written a routine to estimate a complementary log-log model that incorporates a finite-mixture correction, **hshaz**

2) A more general set of routines for Stata is available through the **gllamm** package written by S. Rabe-Hesketh (see http://www.gllamm.org)

f. The aML software package also will estimate models with finite mixture distributions

H. Markov Transition Models for Discrete-Time Processes

1. Notation and data

a. Consider a discrete-time process (exits or transitions observed at discrete intervals)

b. Let $Y_{i,t}$ be a dummy variable that indicates whether person i occupies state 0 or state 1 at time period t

c. In a *Markov model*, $Y_{i,t}$ depends on the past realizations $Y_{i,t-1}$, $Y_{i,t-2}$, …

d. In a *first-order Markov model*, $Y_{i,t}$ only depends on the immediate past realization $Y_{i,t-1}$

e. We can consider the probabilities associated with making different types of transitions

1) let $P_{jk}(t)$ be the probability of transitioning from state j (= 0, 1) at time t–1 to state k (= 0, 1) at time t

2) possible transition probabilities for a first-order Markov model are $P_{01}(t)$, $P_{00}(t)$, $P_{10}(t)$, and $P_{11}(t)$, where $P_{00}(t) = 1 - P_{01}(t)$ and $P_{11}(t) = 1 - P_{10}(t)$

f. Finally, consider a *stationary first-order Markov model*, in which $P_{01}(t) = P_{01}$ and $P_{10}(t) = P_{10}$ for all t

2. It is straightforward to estimate stationary first-order Markov models using standard binary outcome methods

   a. Assume that the transition probabilities are functions of observable characteristics, X_i

   b. Those functions could be consistent with

      1) Logistic distributions (logit models)

      2) Normal distributions (probit models)

      3) Extreme-value distributions (complementary log-log models)

3. Similarly, we could assume that the models are stationary conditional on the observed variables, which would allow us to include time-varying observable variables

4. Models are a restricted version of the discrete-time hazard models (models have no duration dependence)

5. Advantages:

   a. Models are relatively easy to estimate, as they use standard software

   b. Require less data than standard hazard models—do not need the entire history of the process or the duration of the spell

   c. Can use these models with left-censored data

   d. Models are commonly used when there are only a few periods of data (can be estimated using only two observations on the outcome variable)

6. Main disadvantage is the strong assumption on duration dependence
7. If the data are left-censored and the unobserved determinants of the outcomes are correlated, you need to account for initial conditions (i.e., selectivity associated with the first state that you are able to observe)
   a. Formal approach is to calculate the probability of a given initial state given all of the possible earlier transitions; this can be cumbersome
   b. Informal approach is to estimate an approximate model for the initial condition (e.g., simple probit model), allowing for correlations between the unobserved determinants of the initial and subsequent outcomes

I. Repeated events

1. Single spells versus multiple spells
   a. Single spell models are appropriate for describing many types of outcomes such as mortality, transitions to first marriages, initial onset of drinking or sexual activity, etc.
   b. However, many processes are described by repeated spells or repeated periods at risk for the same individual; examples include
      1) Unemployment or employment spells

2) Welfare spells

3) Activity spells (time use episodes)

c. The primary statistical issue that arises with repeated spells is that they are not likely to be completely independent of one another, leading to concerns about

1) Appropriate calculation of standard errors

2) Controls for time-invariant unobserved characteristics

2. Standard errors

a. At a minimum, we would like the calculation of the standard errors to reflect associations across the spells

b. Standard errors for repeated spell models should therefore account for clustering (e.g., using the **robust** or **cluster** options in Stata)

c. These corrections do not change the coefficient estimates, just the coefficient variance/covariance estimates

3. If the source of the correlation across spells is a time-invariant unobserved characteristic, we can use random effects corrections (as in any other panel model)

a. The correction in the repeated spell case is similar to the general correction for unobserved heterogeneity

b. The only difference is that the unobserved term (μ in our earlier models) applies to <u>all</u> of the observations for an individual instead of just a single spell

c. For continuous hazard (**streg**) models in Stata, common random effects can be incorporated using the "shared frailty" option

   1) In **streg**, include the **frailty** option

   2) In addition, include the **shared(***group_id***)** option, where *group_id* is an identifier common to all of the repeated events for a person (or some other group)

d. For discrete-time hazards with normally distributed random effects in Stata

   1) Use the **xtlogit** or **xtcloglog** commands with random effects (the **re** option)

   2) Specify the **i( )** option to use a person-specific identifier instead of a spell-specific identifier

e. Repeated discrete- and continuous-time event history models with normal and finite mixture random effects can also be estimated in aML

J. Multiple outcomes

1. Description of problem

   a. In our analyses so far we have assumed that there are only two states that a person can be in, with the person starting in state 0 and transitioning to state 1

   b. What happens, however, when there are several possible states that a person can transition to

   c. Example from labor economics:

      1) Consider someone who is initially unemployed

    2) The person can transition to employment
    3) But the person can also drop out of the labor force

2. Competing risk (multiple latent risk) framework
    a. Consider a person who starts a spell in state 0
    b. Over the course of that spell, assume that s/he is at risk of transitioning to one of several states
    c. The transition that occurs first is observed
    d. All of the other risks are latent (unobserved)
3. If the exit times are continuous
    a. There is no possibility of multiple transitions occurring at the same instant
    b. The competing risk hazard could be modeled by treating
        1) The transition into the state of interest as an exit event (or failure)
        2) Transitions into any other state as censoring events
        3) Model is the same as the standard hazard, just with a different definition of censoring
4. If the exit times are discrete
    a. There is some possibility of alternative transitions
    b. Competing risk should be modeled using some type of unordered multinomial choice specification, such as multinomial logit
    c. If multinomial logit is used, the set up of the data is very similar to the standard approach

d. Suppose that transition to one of K possible states occurs at duration t; construct observations so that
   1) Observations for first $t-1$ periods have outcomes $Y_j = 0$ ($j = 1, t-1$)
   2) Observation for the t^{th} period has outcome $Y_t = k$, where $k = 1, 2, \ldots K$ depending on the type of transition
e. If the observation is right-censored at t, construct a sequence with all zeros
f. Each outcome would also be accompanied by measures of the observed characteristics in each time period X_j, for $j = 1, t$
g. With these data, you could then run a standard multinomial logit model in which the Y_j values are the dependent variables and the X_j are the independent variables

References:

Allison, Paul. "Discrete-Time Methods for the Analysis of Event Histories." *Sociological Methodology* 13 (1982), 61-98.

Greene, William. *Econometric Analysis* 5th edition. Upper Saddle River, NJ: Prentice Hall, 2003, Chapter 22.

Heckman, James J., and Burton Singer. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (March 1984), 271-320.

Lillard, Lee. "Simultaneous equations for hazards: Marriage duration and fertility timing." *Journal of Econometrics* 56 (1993): 189-217.

Moffitt, Robert A., and Michael S. Rendall. "Cohort Trends in the Lifetime Distribution of Female Family Headship in the United States, 1968-1985," *Demography* 32:3 (August 1995), 407-24.

Ribar, David C. "Transitions From Welfare and the Employment Prospects of Low-skill Workers," *Southern Economic Journal* 71:3 (January 2005), 514-33.