Lecture notes from workshops on
Multilevel Discrete-time Event History Analysis

Fiona Steele
Centre for Multilevel Modelling
Graduate School of Education
University of Bristol

Course Outline

- Discrete-time methods for modelling time to a single event
- Multilevel models for recurrent events and unobserved heterogeneity
- Modelling transitions between multiple states
- Modelling competing risks
- Multiprocess models for correlated histories

Discrete-time Methods for Modelling the Time to a Single Event

What is Event History Analysis?

Methods for analysis of length of time until the occurrence of some event. The dependent variable is the duration until event occurrence.

EHA also known as:

- Survival analysis (particularly in biostatistics and when event is not repeatable)
- Duration analysis
- Hazard modelling

Examples of Applications

- **Education** – time to leaving full-time education (from end of compulsory education); time to exit from teaching profession

- **Economics** – duration of an episode of unemployment or employment

- **Demography** – time to first birth (from when?); time to first marriage; time to divorce

- **Psychology** – duration to response to some stimulus

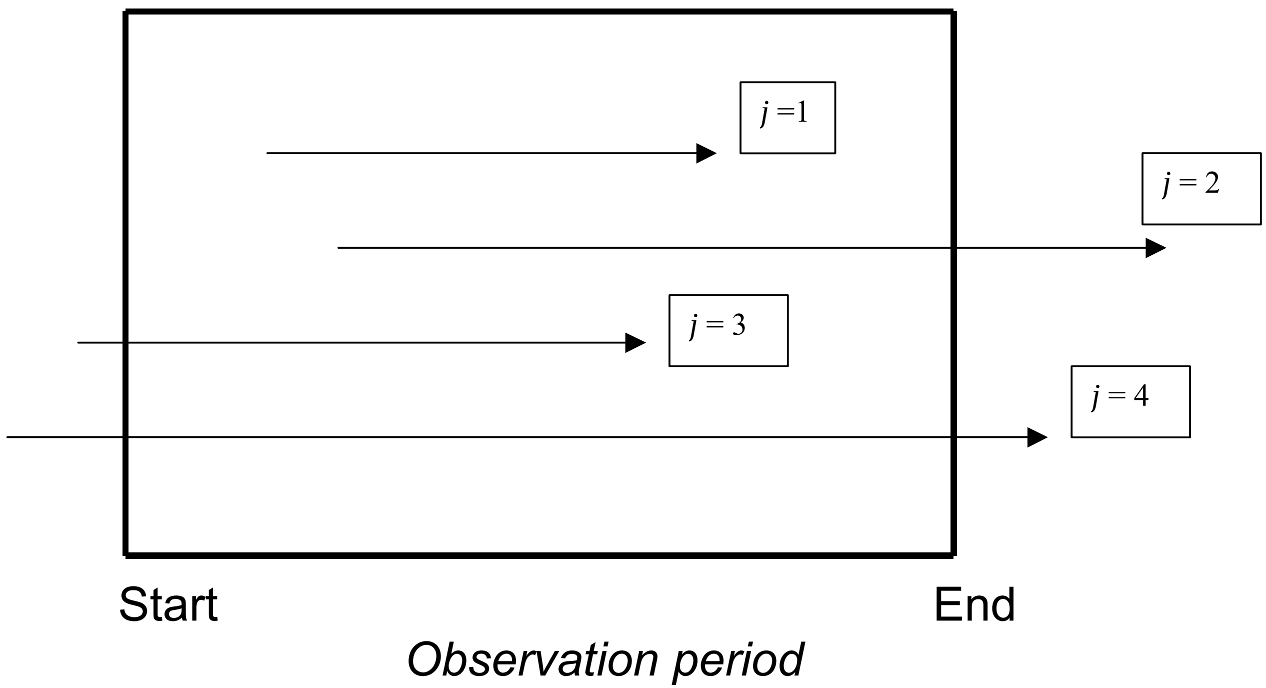
Special Features of Event History Data

- Durations are always positive and their distribution is often skewed

- Censoring – there are usually people who have not yet experienced the event when we observe them

- Time-varying covariates – the values of some covariates may change over time

Censoring (1)



Censoring (2)

Arrowhead indicates time that event occurs.

j = 1 start and end time known
j = 2 end time outside observation period, i.e. *right-censored*
j = 3 start time outside observation period, i.e. *left-truncated*
j = 4 start and end time outside observation period

Right-censoring is most common form of incomplete observation, and is straightforward to deal with in EHA.

Right-censoring: Non-informative Assumption

Excluding right-censored observations leads to bias and may drastically reduce sample size. In EHA we retain these observations and usually make the assumption that censoring is non-informative, i.e. event times are independent of censoring mechanism (like the 'missing at random' assumption).

Assume individuals are not selectively withdrawn from the sample because they are more or less likely to experience an event. May be questionable in experimental research, e.g. if more susceptible individuals were selectively withdrawn from a 'treatment' group.

Event Times and Censoring Times

Denote the event time (also known as **duration, failure, or survival time**) by the random variable T.

t_j event time for individual j

δ_j censoring indicator
=1 if uncensored (i.e. observed to have event)
=0 if censored

But for right-censored case, we do not observe t_j. We only observe the time at which they were censored, c_j.

Our dependent variable is $y_j = \min(t_j, c_j)$.

Our observed data are (y_j, δ_j).

The Discrete-time Approach

Event times measured in discrete intervals t=1, 2, 3, . . . (e.g. months, years).

Can think of event history as a series of independent success/failure trials. In each interval t we observe a binary response indicating whether an event has occurred.

Main Advantages of the Discrete-time Approach

- Events times often measured in discrete-time units, particularly when collected retrospectively

- Allows proportional hazards assumption to be tested straightforwardly. Straightforward to allow for non-proportional hazards.

- Analysis straightforward as we can use models for discrete response data – important for more complex data structures and processes

Disadvantages of the Discrete-time Approach

- Data must first be restructured so that for each individual we have a sequence of observations, one for each time interval until event occurrence or censoring.

- If observation period is long relative to the width of the time intervals in which durations are measured, the dataset may become very large.

Discrete-time Hazard Function

In EHA, interest is usually focused on the **hazard function**, *h*(*t*), and how it depends on covariates.

The discrete-time hazard function *h*(*t*) is the probability of having at event during interval *t*, given no earlier occurrence:

$$h(t) = \Pr(T = t \mid T \geq t),$$

where T is the event time.

Discrete-time hazard often denoted by h_t

Non-parametric Estimation of h(t)

$r(t)$ is no. at ‘risk’ of experiencing event at start of interval t

$d(t)$ is no. of events observed during interval t

$w(t)$ is no. of censored cases in interval t

The life table (or actuarial estimator) of $h(t)$ is

$$\hat{h}(t) = \frac{d(t)}{r(t) - w(t)/2}$$

Note: assume censoring times are spread uniformly across interval t. Some estimators ignore censored cases.

Discrete-time Survivor Function

The **survivor function**, S(t), is the probability that an event has not occurred before time t:

$$S(t) = \Pr(T \geq t)$$

The probability that an event occurs before time t is F(t) = Pr(T < t) = 1 – S(t). F(t) is the **cumulative density function**.

Estimation of S(t)

Estimator of survivor function for interval *t* is

$$\begin{aligned}\hat{S}(t) &= [1-\hat{h}(1)]\times[1-\hat{h}(2)]\ldots\times[1-\hat{h}(t-1)] \\ &= \hat{S}(t-1)\times[1-\hat{h}(t-1)]\end{aligned}$$

Note: Sometimes survivor function defined as S(*t*)=Pr(*T*>*t*).

Restructuring Data for a Discrete-time Analysis: Individual-based File

E.g. records for 2 individuals

INDIVIDUAL (j)	DURATION (t_j)	CENSOR (δ_j)	AGE (x_j)
1	5	0	20
2	3	1	35

CENSOR=1 if uncensored and 0 if censored.

Restructuring Data for a Discrete-time Analysis: Discrete-time (Person-period) File

j	t	$y_j(t)$	x_j
1	1	0	20
1	2	0	20
1	3	0	20
1	4	0	20
1	5	0	20
2	1	0	35
2	2	0	35
2	3	1	35

$y_j(t)$= 1 if event occurs to individual j at time t
= 0 if event has not occurred

The Discrete-time Logit Model (1)

The response variable for a discrete-time model is the binary indicator of event occurrence $y_j(t)$.

The **hazard function** may be written as

$$h_j(t) = \Pr(y_j(t) = 1 \mid y_j(t-1) = 0)$$

The Discrete-time Logit Model (2)

We can fit a logit regression model of the form:

$$\text{logit}\,[h_j(t)] = \log\left[\frac{h_j(t)}{1-h_j(t)}\right] = \alpha(t) + \beta x_j(t) \qquad (*)$$

The covariates $x_j(t)$ can be constant over time or time-varying.

$\alpha(t)$ is some function of time, called the logit of the **baseline hazard function**. This needs to be specified.

Modelling the Time-dependency of the Hazard (1)

Changes in $h(t)$ over time are captured by $\alpha(t)$. This might be a **linear** or **quadratic** function.

Linear: $$\alpha(t) = \alpha_0 + \alpha_1 t$$

Quadratic: $$\alpha(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

Modelling the Time-dependency of the Hazard (2)

In the most flexible model, time is treated as a categorical variable with a category for each time interval, leading to a step function:

$$\alpha(t) = \alpha_1 D_1 + \alpha_2 D_2 + \ldots + \alpha_q D_q$$

where D_1, D_2, . . ., D_q are dummies for time intervals t=1, 2, . . ., q, and q is the maximum observed event time. (Alternatively we can choose one time interval as the reference and fit an overall intercept term.)

If q is very large, categories can be grouped together – piecewise constant hazard model.

The Proportional Hazards Assumption

Model (*) assumes that the effects of covariates $x(t)$ are constant over time. This is known as the **proportional hazards** assumption. (Strictly it is the ***odds*** that are assumed proportional as we are fitting a logit model.)

We can relax this assumption by introducing interactions between $x(t)$ and $\alpha(t)$.

Grouping Time Intervals

When we move to more complex models, a potential problem with the discrete-time approach is that creating one record per discrete time unit may lead to a large dataset.

It may be possible to group time intervals, e.g. using 6-month rather than monthly intervals. In doing so, we have to assume that the hazard and values of covariates are constant within the grouped intervals.

Analysing Grouped Time Intervals

If we have grouped time intervals, we need to allow for different lengths of exposure time within these intervals.

For example, for any 6-month interval, some individuals will have the event or be censored after the first month while others will be exposed for the full 6 months. Denote by $n_j(t)$ the exposure time for individual j in grouped interval t.

We then define a new response $y^*_j(t)=y_j(t)/\ n_j(t)$ and declare $n_j(t)$ as a denominator for this proportion.

<u>Note</u>: intervals do not need to be the same width.

Example of Grouped Time Intervals

Suppose an individual is observed to have an event during the 17th month, and we wish to group durations into 6-month intervals (*t*).

j	t	$n_j(t)$	$y_j(t)$	$y^*_j(t)$
1	1	6	0	0
1	2	6	0	0
1	3	5	1	0.2

Example: Singer & Willett (1993)

Longitudinal data following career paths of 3,941 special educators in Michigan, hired between 1972 and 1978.

Event of interest is stopping teaching by 1985 (end of observation period). So maximum duration is 12 years. Minimum censoring time is 7 years.

Hazard and Survivor Functions

t	*r*(*t*)	*d*(*t*)	*w*(*t*)	*h*(*t*)	*S*(*t*)
1	3941	456	0	0.116	1
2	3485	384	0	0.110	0.884
3	3101	359	0	0.116	0.787
4	2742	295	0	0.108	0.696
5	2447	218	0	0.089	0.621
6	2229	184	0	0.083	0.566
7	2045	123	280	0.065	0.519
8	1642	79	307	0.053	0.486
9	1256	53	255	0.047	0.460
10	948	35	265	0.043	0.438
11	648	16	241	0.030	0.419
12	391	5	386	0.025	0.407

Examples of Interpretation

10.8% of teachers who were still teaching at the start of their 4th year, left during their 4th year [$h(4)$].

40.7% of teachers were still teaching at the start of their 12th year [$S(12)$].

Discrete-time Logit Model

The following results are from fitting a model of the form

$$\text{logit}[h(t)] = \alpha(t) + \beta \text{ FEMALE}$$

where $\alpha(t) = \alpha_1 D_1 + \alpha_2 D_2 + \ldots + \alpha_{12} D_{12}$

$D_1, D_2, \ldots, D_{12}$ are dummies for years 1, 2, . . . 12
(no overall intercept)

FEMALE is a dummy for sex.

Results from Fitting Logit Model

	Estimate (SE)
D_1	-2.41 (0.08)
D_2	-2.47 (0.08)
D_3	-2.41 (0.08)
D_4	-2.49 (0.09)
D_5	-2.70 (0.09)
D_6	-2.78 (0.10)
D_7	-3.12 (0.11)
D_8	-3.35 (0.13)
D_9	-3.48 (0.15)
D_{10}	-3.62 (0.18)
D_{11}	-4.03 (0.26)
D_{12}	-4.69 (0.45)
FEMALE	0.44 (0.07)

Unobserved Heterogeneity

Introduction

Some individuals will be more at risk of experiencing an event than others, and it is unlikely that the reasons for this variability in the hazard will be fully captured by covariates.

The presence of unobserved (or unobservable) individual-specific risk factors leads to **unobserved heterogeneity** in the hazard.

Unobserved heterogeneity is also referred to as **frailty**, particularly in biostatistics (more ‘frail’ individuals have a higher mortality hazard).

Consequences of Unobserved Heterogeneity

If there are individual-specific unobserved factors that affect the hazard, the observed form of the hazard function at the aggregate population level will tend to be different from those at the individual level.

Even if the hazards of individuals in a population are constant over time the aggregate population hazard may be time-dependent, typically decreasing. This may be explained by a **selection effect** operating on individuals.

Selection Effect

If a population is heterogeneous in its susceptibility to experiencing an event, high risk individuals will tend to have the event first, leaving lower risk individuals in the population.

Therefore as t increases the population is increasingly depleted of those individuals most likely to experience the event, leading to a decrease in the population hazard.

Because of this selection, we may see a decrease in the population hazard even if individual hazards are constant (or even increasing).

Impact of Unobserved Heterogeneity on Parameter Estimates

If unobserved heterogeneity is incorrectly ignored:

(i) A positive duration dependence will be underestimated, while a negative duration dependence will be overestimated.

(ii) The magnitude of regression coefficients will be underestimated. BUT note that estimates from a frailty model have a different interpretation – see later.

Allowing for Unobserved Heterogeneity in a Discrete-Time Model

We can introduce a random effect which represents individual-specific unobservables:

$$\text{logit}[h_j(t)] = \alpha(t) + \beta x_j(t) + u_j$$

Usually assume $u_j \sim N(0, \sigma_u^2)$

σ_u^2 represents unobserved heterogeneity or frailty.

Interpretation of Coefficients from a Frailty Model

Suppose we have a continuous covariate x, with coefficient β.

In model without frailty, $\exp(\beta)$ is an odds ratio. It compares the odds of an event for two randomly selected individuals with x-values 1 unit apart (and the same values for other covariates in the model). $\exp(\beta)$ is the population averaged effect of x.

In model with frailty, $\exp(\beta)$ is an odds ratio only when the random effect is held constant, i.e. if we are comparing two hypothetical individuals with the same random effect value. $\exp(\beta)$ is the individual-specific effect of x.

Example

Time to first partnership (Practical Exercise 1)

2 dichotomous covariates:

FEMALE

FULLTIME (in full-time education), time-varying

Consider models with and without frailty.

Results from Fitting Models Without (1) and With (2) Unobserved Heterogeneity

	Model 1 Est. (SE)	**Model 2** Est. (SE)
t	0.04 (0.01)	0.11 (0.04)
FEMALE	0.43 (0.10)	0.59 (0.15)
FULLTIME	-1.51 (0.18)	-1.56 (0.19)
σ_u	-	0.72 (0.22)

LR test statistic for comparison of models = 6.09 on 1 d.f.

Time to First Partnership: Interpretation of Coefficients from the Frailty Model

For a given individual, the odds of entering a partnership at age t when in full-time education are exp(-1.56)=0.21 times the odds when not in full-time education. This interpretation is useful because FULLTIME is time-varying within an individual.

For 2 individuals with the same random effect value, the odds are exp(0.59)=1.8 times higher for a woman than for a man. This interpretation is less useful. We could, however, obtain a population average effect of sex as follows.

Obtaining Population Average Effects from a Frailty Model: Predicted Hazards

Compute $h(t)$ for $x(t)=x^*(t)$:

$$h^*(t) = \frac{\exp[\alpha(t) + \beta x^*(t) + u]}{1 + \exp[\alpha(t) + \beta x^*(t) + u]}$$

What value for u? $u=0$, i.e. the mean?

Because of nonlinearity of logistic transformation, $h^*(t)$ at $u=0$ is **not** equal to the mean $h^*(t)$. It is actually the median.

Solutions: integrate out u, or simulate distribution of u.

Predicted Hazards by Simulation

(i) Generate M values for random effect *u* from $N(0, \hat{\sigma}_u^2)$: $u_{(1)}, u_{(2)}, \ldots, u_{(M)}$

(ii) For $m = 1,\ldots,M$ compute, for $\mathbf{x}(t) = \mathbf{x}^*(t)$:

$$h^*_{(m)}(t) = \frac{\exp[\alpha(t) + \beta x^*(t) + u_{(m)}]}{1 + \exp[\alpha(t) + \beta x^*(t) + u_{(m)}]}$$

(iii) Mean (population average) hazard:

$$h^*(t) = \frac{1}{M} \sum_m h^*_{(m)}(t)$$

Population Average Odds Ratio from Predicted Hazards

Suppose interested in obtaining the odds ratio for females relative to males.

Mean of M hazards (from simulation)	FEMALE (x=x^*)
$h_1^*(t)$	1
$h_0^*(t)$	0

$$OR_{PA} = \frac{h_1^*(t)/[1-h_1^*(t)]}{h_0^*(t)/[1-h_0^*(t)]}$$

Estimation of Model with Unobserved Heterogeneity

- Estimate using any software for random effects (multilevel) logit models. Options include:
  - SAS (`proc nlmixed`), Stata (`xtlogit`), aML (all use numerical quadrature)
  - MLwiN (quasi-likelihood, MCMC)
  - WinBUGS (MCMC)

Markov Chain Monte Carlo (MCMC) Estimation (1)

- Bayesian formulation that puts priors on parameters and generates a (correlated) chain of sample draws from the posterior distribution of the parameters.

- WinBUGS is the most flexible software – *MLwiN* tailored to multilevel data and uses ‘good’ starting values.

- Consider the 2-level variance components model for Normal response data:

$$y_{ij} = \beta_0 + \beta_{1ij} x_{ij} + u_j + e_{ij}, \quad u \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2)$$

MCMC Estimation (2)

The basic idea is that we require the joint (posterior) distribution of the parameters given the data *y* and (independent) prior distributions for each of the parameters.

A ‘prior’ represents information about a parameter before we collect data. For example a ‘diffuse’ prior is

$$\beta \sim N(0,10000) \cong U(-\infty,\infty)$$

This is then combined with the data (via the likelihood for the data) to produce a ‘posterior’ *distribution* for the parameter.

MCMC works by drawing a random sample of sets of parameter values, say 5,000 sets, and basing inference on these ‘chains’.

MCMC Estimation (3)

At iteration *n* we want to sample from the posterior distribution of each parameter in turn. If we can write down an analytical expression for the posterior – as is the case for Normal models - then Gibbs sampling is available and efficient.

Otherwise we need an alternative – *MLwiN* uses Metropolis-Hastings sampling, e.g. for binary response models.

MCMC Estimation (4)

The MCMC procedure for a 2-level binary response model is as follows:

1. Choose starting values (quasi-likelihood)

2. Sample a new set of fixed effects given the current values of all the other parameters and the data

3. Sample a new set of random effects given other parameters and data

4. Sample new random effect variance

…..etc.

Modelling Recurrent Events

Examples of Recurrent Events

Many types of event can occur more than once to an individual. Define an episode as the time between the start of the 'risk' period and the occurrence of an event or censoring.

Examples

Employment episode: duration from starting a new job to leaving that job.

Marriage episode: duration of a marriage.

Problem with Analysing Recurrent Events

We cannot assume that the durations of episodes from the same individual are independent.

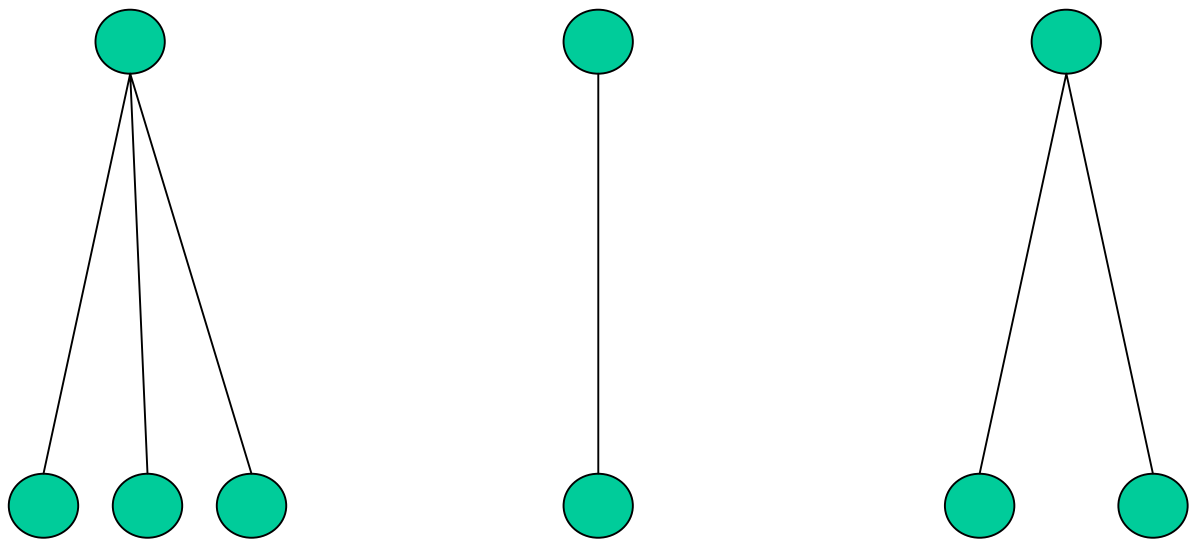
There may be unobserved individual-specific factors (i.e. constant across episodes) which affect the hazard of an event for <u>all</u> episodes. Presence of such unobservables, and failure to account for them in the model, will lead to correlation between the durations of episodes from the same individual.

Hierarchical Data Structure

Recurrent events lead to a two-level hierarchical structure.

Level 2: Individuals

Level 1: Episodes



Simple Two-level Model for Recurrent Events (1)

$$\text{logit}[h_{ij}(t)] = \alpha(t) + \beta x_{ij}(t) + u_j$$

$h_{ij}(t)$ is hazard of event in time interval *t* during episode *i* of individual *j*

$x_{ij}(t)$ are covariates which might be time-varying or defined at the episode or individual level

u_j random effect representing unobserved characteristics of individual *j* – **shared 'frailty'** (common to all episodes)

Assume $u_j \sim N(0, \sigma_u^2)$

Simple Two-level Model for Recurrent Events (2)

- The model for recurrent events is essentially the same as the model for unobserved heterogeneity, and is therefore estimated in exactly the same way.

- Recurrent events allow better identification of the random effect variance.

- The expansion of data to discrete-time format is carried out for each episode within an individual.

Episode-specific Effects

We can allow the duration and covariate effects to vary between episodes. E.g. we might expect factors affecting the timing of the first event to differ from those affecting timing of subsequent events (or the same factors to have different effects).

Include dummy variable(s) for order of the event and interact with *t* and covariates.

Example

Repeated birth intervals for Hutterite women, a natural fertility (no contraceptive use) population in North America.

Data on 944 birth intervals from 159 women. Interval is duration between a birth and conception of next child.

Only closed intervals (i.e. ending in a conception before the survey). Long open intervals may be due to primary or secondary sterility. Therefore there is no censoring.

Duration and Covariates

Duration of a birth interval is represented by a categorical variable (in a piecewise constant hazards model):

MONTH	Month of exposure to risk of conception [<6, 6-11, 12-23, 24-35, 36+ (ref.)]

Consider one covariate:

AGE	Age at start of birth interval (years)

Results

	Coeff.	(SE)
Const	0.38	(0.39)
MONTH		
<6	-0.96	(0.30)
6-11	-0.21	(0.30)
12-23	0.12	(0.30)
24-35	-0.24	(0.36)
AGE	-0.07	(0.01)
σ_u^2	0.31	(0.06)

Random Coefficient Models

The model we have considered so far assumes that unobserved heterogeneity is constant. Using multilevel modelling terminology, this model is a random intercept model. The form of the hazard is assumed the same across individuals, but is shifted up or down by an amount u_j on the logit scale. The duration and covariate effects are assumed to be the same for each individual.

To test these assumptions, we can consider random coefficient models (called random slope models for continuous covariates with linear effects). Note, however, that there may be insufficient information to estimate random coefficients if the number of individuals with recurrent events is small.

Random Coefficient for AGE

Suppose we suspect that the effect of AGE varies across women. We explore this using a random coefficient model:

$$\text{logit}[h_{ij}(t)] = \alpha(t) + \beta_j \, AGE_{ij} + u_{0j}$$

where $\beta_j = \beta + u_{1j}$

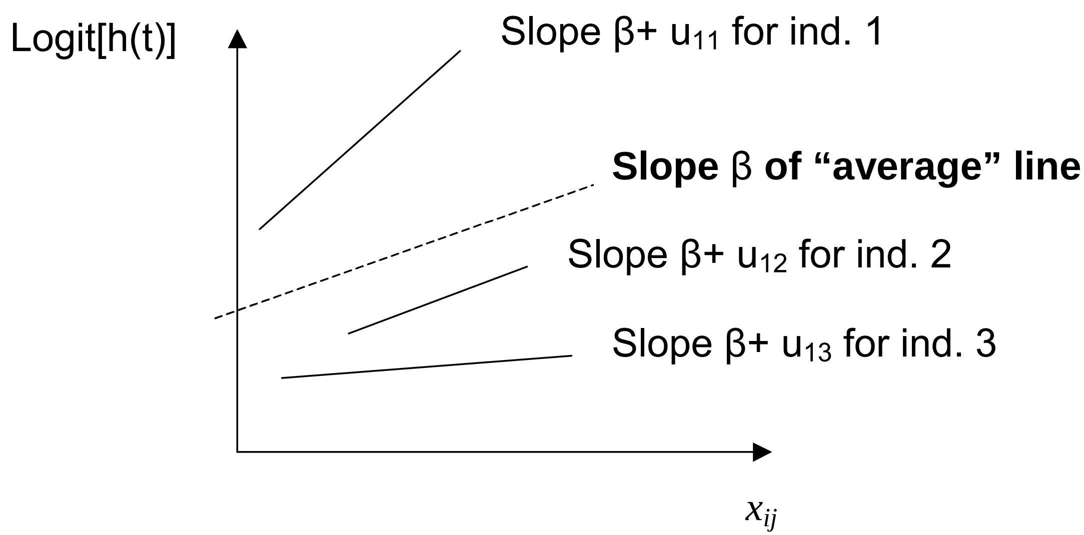
Also written as $$\text{logit}[h_{ij}(t)] = \alpha(t) + \beta \, AGE_{ij} + u_{0j} + AGE_{ij} u_{1j}$$

Assume $u_{0j} \sim N(0, \sigma^2_{u0}), \qquad u_{1j} \sim N(0, \sigma^2_{u1})$

$$\text{Cov}(u_{0j}, u_{1j}) = \sigma_{u01}$$

Graphical Representation of a Random Coefficient Model

For any time interval *t*:



Unobserved Heterogeneity as a Function of AGE

In the model with a random coefficient for AGE, the unobserved heterogeneity between women is:

$$\mathrm{Var}(u_{0j} + AGE_{ij}u_{1j})$$

$$= Var(u_{0j}) + 2Cov(u_{0j}, u_{1j})AGE_{ij} + Var(u_{1j})AGE_{ij}^2$$

$$= \sigma_{u0}^2 + 2\sigma_{u01}AGE_{ij} + \sigma_{u1}^2 AGE_{ij}^2$$

i.e. a quadratic function in AGE.

Results from Random Coefficient Model (AGE centred around mean)

	Coeff.	(SE)
Const	-1.55	(0.30)
MONTH		
<6	-0.99	(0.31)
6-11	-0.21	(0.30)
12-23	0.14	(0.30)
24-35	-0.21	(0.37)
AGE	-0.07	(0.01)
σ^2_{u0}	0.29	(0.85)
σ_{u01}	-0.007	(0.005)
σ^2_{u1}	0.001	(0.001)

Testing Significance of Random Coefficient

2 additional parameters introduced to random intercept model:

σ_{u1}^2 and σ_{u01}

Test the null hypothesis that $H_0 : \sigma_{u1}^2 = \sigma_{u01} = 0$

The (approximate) Wald test statistic is 2.84 on 2 d.f.

So fail to reject null and conclude that the effect of AGE is constant across women.

Interpretation of Random Coefficient for AGE

In practice, we would revert to the random intercept model after finding little evidence of a random coefficient. But, for illustration, we consider the interpretation of the random coefficient model.

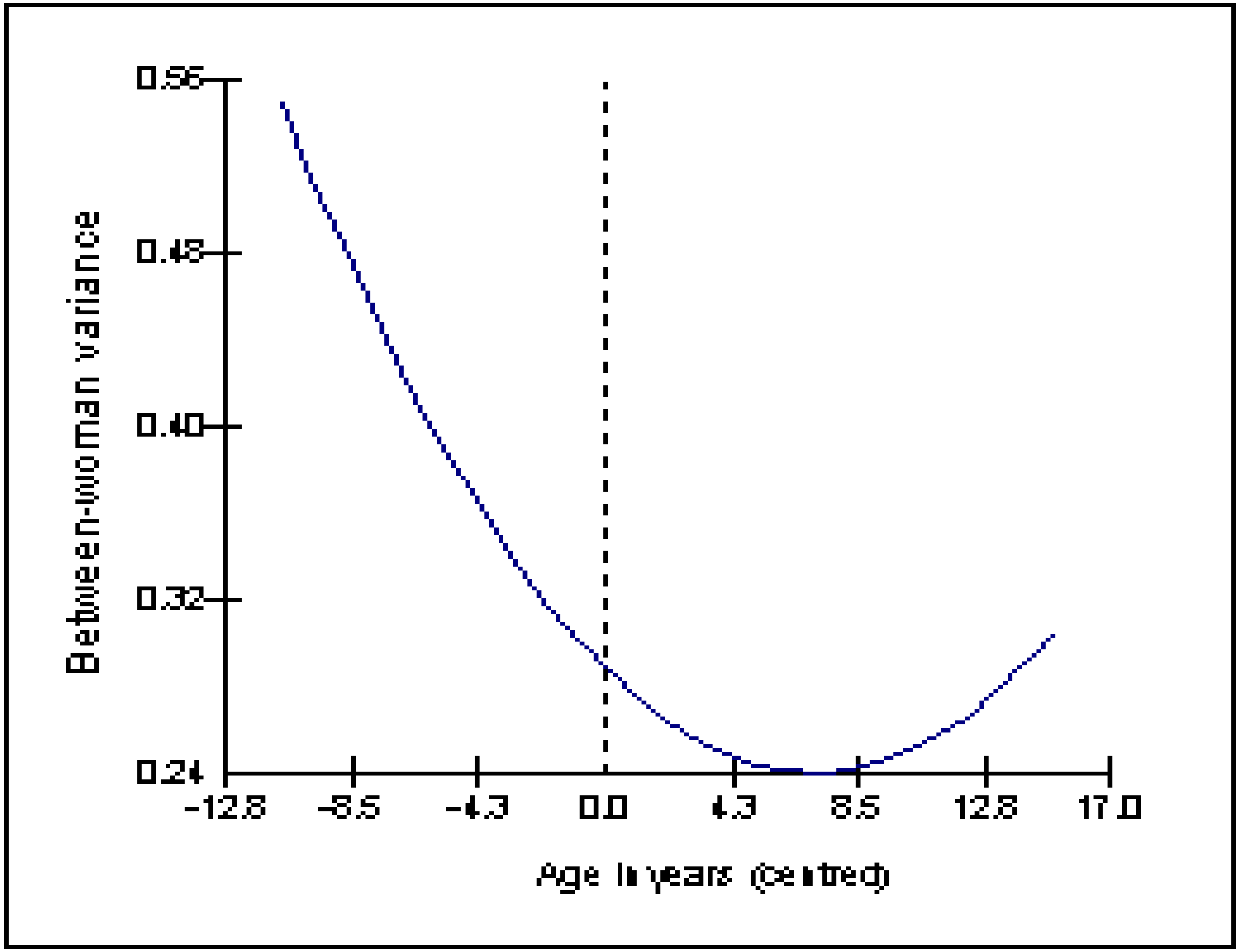
The between-woman variance in the logit-hazard of conception (after accounting for duration effects) is:

$$\mathrm{Var}(u_{0j} + AGE_{ij}u_{1j})$$

$$= 0.29 - 0.014\,AGE_{ij} + 0.001\,AGE_{ij}^{2}$$

We can plot the between-woman variance as a function of AGE.

Between-woman Unobserved Heterogeneity as a Function of AGE



Multiple States

States in Event Histories

In the models we have considered so far, there is a single event of interest. We model the duration to this event from the point at which an individual becomes "at risk". We can think of this as the duration spent in the same **state**.

E.g. in a study of marital dissolution we model the duration in the **marriage state**.

In a study of birth intervals we model the duration spent in the **non-pregnant state** (up to age 45 say).

Multiple States

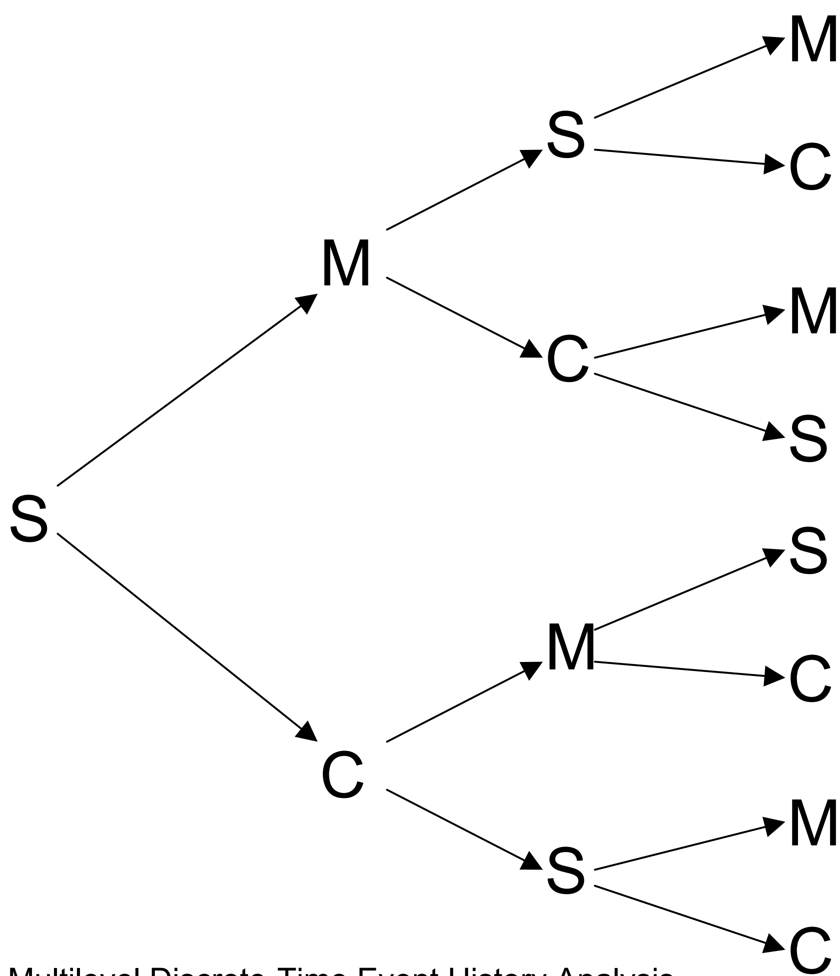
Usually individuals will move in and out of different states over time, and we wish to model these transitions.

Examples

Partnership states: marriage, cohabitation, single (unpartnered)

Employment states: employed, unemployed, out of the labour market

Possible Partnership Transitions
(S=single, M=marriage, C=cohabitation)



etc.

Note that censoring can occur in any state, at which point no further transitions are observed.

Example: Transitions Between Partnership States

For simplicity, group marriage and cohabitation into one ‘partnership’ state (P), and model transitions between this state and the ‘single’ state (S).

Need to estimate two equations: (1) duration spent in P, where event is moving to S; (2) duration spent in S, where event is moving to P.

Model for Transitions Between Partnership and Single States

(Adapted from Goldstein et al. 2004, *Understanding Statistics*)

Transitions from P to S (may be multiple transitions per individual):

$$\text{logit}[h_{ij}^{P}(t)] = \alpha^{P}(t) + \beta^{P} x_{ij}^{P}(t) + u_{j}^{P}$$

Transitions from S to P (actually fit separate eq. for first S-P transition):

$$\text{logit}[h_{ij}^{S}(t)] = \alpha^{S}(t) + \beta^{S} x_{ij}^{S}(t) + u_{j}^{S}$$

Allow correlation between u_{j}^{P} and u_{j}^{S}.

Why Allow Correlation Between Random Effects Across States?

There may be time-invariant individual-specific unobservables that affect each type of transition.

E.g. individuals with a strong desire to be in a partnership might have a low hazard of moving from P to S, and a high hazard of moving from S to P, i.e. a tendency towards long partnerships and short periods in the single state.

This would lead to a negative random effect correlation.

Random Effect Covariance Matrix
(from Goldstein et al. 2004)

	P	S
P	0.462 (0.061)	
S	-0.313 (0.095) -0.524	0.773 (0.141)

Source: National Child Development Study (NCDS), age 16-33
Note: Standard errors in parentheses; correlation estimate.

Interpretation of Random Effect Correlation (Men only)

The estimated correlation of –0.524 implies that there is a moderate negative association between the duration in a partnership and the duration spent without a partner.

We can, tentatively, classify men as long partnership/ short-time single or short partnership/ long-time single.

Estimation of a Multiple State Model

- Specify a single equation model with dummy variables for each state. Interact dummies with duration and covariates to obtain state-specific duration and covariate effects.

- Allow coefficient of each dummy to vary and covary randomly across individuals.

Data Structure I

Start with an episode-based file, e.g.

j	i	$State_{ij}$	t_{ij}	δ_{ij}	Age_{ij}
1	1	S	3	1	16
1	2	P	2	0	19

Notes: (1) t in years; (2) δ_{ij} =1 if uncensored, 0 if censored; (3) age, in years, at start of episode.

Data Structure II

Convert to discrete-time format:

t	$y_{ij}(t)$	S_{ij}	P_{ij}	S_{ij}*Age_{ij}	P_{ij}*Age_{ij}
1	0	1	0	16	0
2	0	1	0	16	0
3	1	1	0	16	0
1	0	0	1	0	19
2	0	0	1	0	19

S_{ij} dummy for Single, P_{ij} dummy for Partnership.

Model Specification

- Specify a 2-level logit model. Include S_{ij}, P_{ij}, S_{ij}*Age_{ij} and P_{ij}*Age_{ij} as explanatory variables. Also include interactions between P_{ij}, S_{ij} and some function of duration t.

- Coefficients of S_{ij} and S_{ij}*Age_{ij} are intercept and effect of age on log-odds of forming a partnership. Coefficients of P_{ij} and P_{ij}*Age_{ij} are intercept and effect of age on log-odds of dissolving a partnership.

- Allow coefficient of S_{ij} and P_{ij} to vary randomly across individuals.

Competing Risks

Examples

There will often be more than one way of (or reason for) exiting a particular state; we call these "competing risks". We assume these are mutually exclusive.

Event	**Competing risks**
Death	Different causes
End of employment spell	Sacked, redundancy, switch job, out of labour market
Partnership formation	Marriage or cohabitation
Contraceptive discontinuation	Different reasons, e.g. failure, side effects

Cause-Specific Hazard Functions

The hazard is defined as for single types of event, but now we have one for each competing risk.

Suppose there are R competing risks, then the hazard of event type *r* at time *t* is:

$$h^{(r)}(t) = \Pr(\text{event of type } r \text{ at time } t \mid T \geq t)$$

The hazard that no event *of any type* occurs at *t* (given 'survival' to *t*) is:

$$h^{(0)}(t) = 1 - \sum_{r=1}^{R} h^{(r)}(t)$$

Calculation of the Survivor Function

The probability of survival is the probability that no event *of any type* occurs before time *t*:

$$S(t) = h^{(0)}(1) \times h^{(0)}(2) \times \ldots \times h^{(0)}(t-1)$$

Modelling Approaches I

In discrete time, we can distinguish between two main modelling approaches:

(1) Model the cause-specific hazards simultaneously using a multinomial logistic model

E.g. for partnership formation, estimate 2 contrasts: marriage vs. single (“no event”), cohabitation vs. single. (The remaining contrast, marriage vs. cohabitation, may be estimated from the other two.)

c.f. Multiple decrement life table

Modelling Approaches II

(2) Model each competing risk separately, treating all other events as censored

E.g. for partnership formation, estimate 2 separate binary logistic models:

(i) Model hazard of marriage, treating transitions to cohabitation as censored;

(ii) Model hazard of cohabitation, treating transitions to marriage as censored.

c.f. Associated single decrement life table

Modelling Approaches III

- Rates estimated using approach (2) represent the underlying (or theoretical) risk of a particular event in the absence of all other risks.

E.g. “What would be the risk of dying from cancer if there were no other causes of death?”

- Useful for comparing rates across time or populations.

E.g. Comparison of death rates due to cancer in 1900 and 1990, adjusting for changes in life expectancy.

Data Structure for Multinomial Model

For each discrete time interval *t* define a multinomial response, $y_{ij}(t)$, indicating occurrence and type of event.

The response categories are 0 (“no event”), 1, 2, . . ., R.

Note: Further data restructuring needed to fit model in MLwiN – see practical.

The Multinomial Logit Model
(Multilevel Version for Recurrent Episodes)

R equations contrasting event type *r* with “no event”.

$$\log\left(\frac{h_{ij}^{(r)}(t)}{h_{ij}^{(0)}(t)}\right) = \alpha^{(r)}(t) + \beta^{(r)} x_{ij}^{(r)}(t) + u_j^{(r)}$$

where $(u_j^{(1)}, u_j^{(2)}, \ldots, u_j^{(R)})$ ~ multivariate normal, variance Ω_u

Random effects are correlated to allow for shared unobserved risk factors.

Estimation of Cause-Specific Hazards

Usual formula for calculating predicted probabilities from a multinomial logit model:

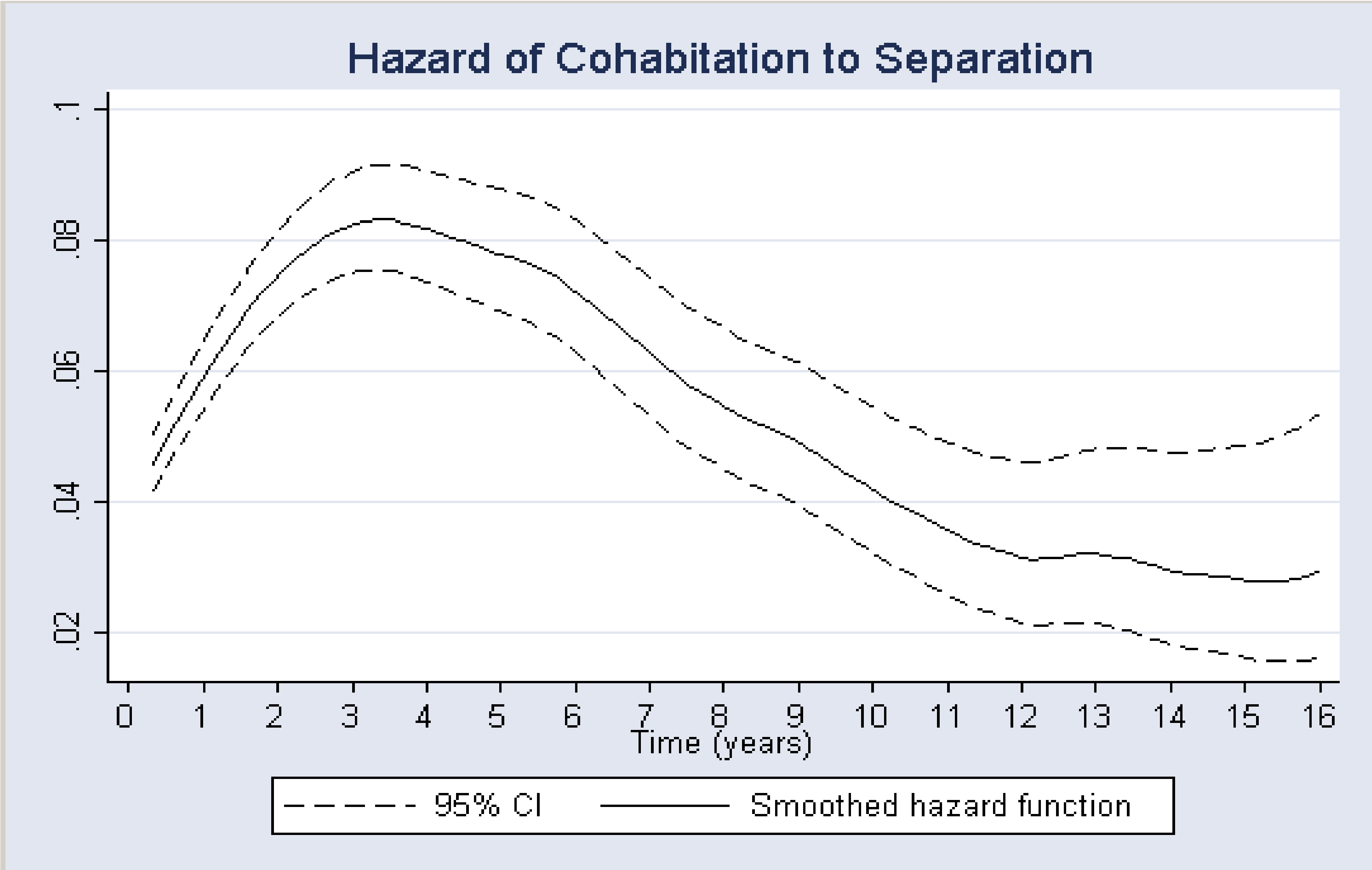
$$\hat{h}^{(r)}(t) = \frac{\exp(\hat{\alpha}^{(r)}(t) + \hat{\beta}^{(r)} x^{(r)}(t) + u^{(r)})}{1 + \sum_{r=1}^{R} \exp(\hat{\alpha}^{(r)}(t) + \hat{\beta}^{(r)} x^{(r)}(t) + u^{(r)})}$$

Need to substitute some value for $u^{(r)}$

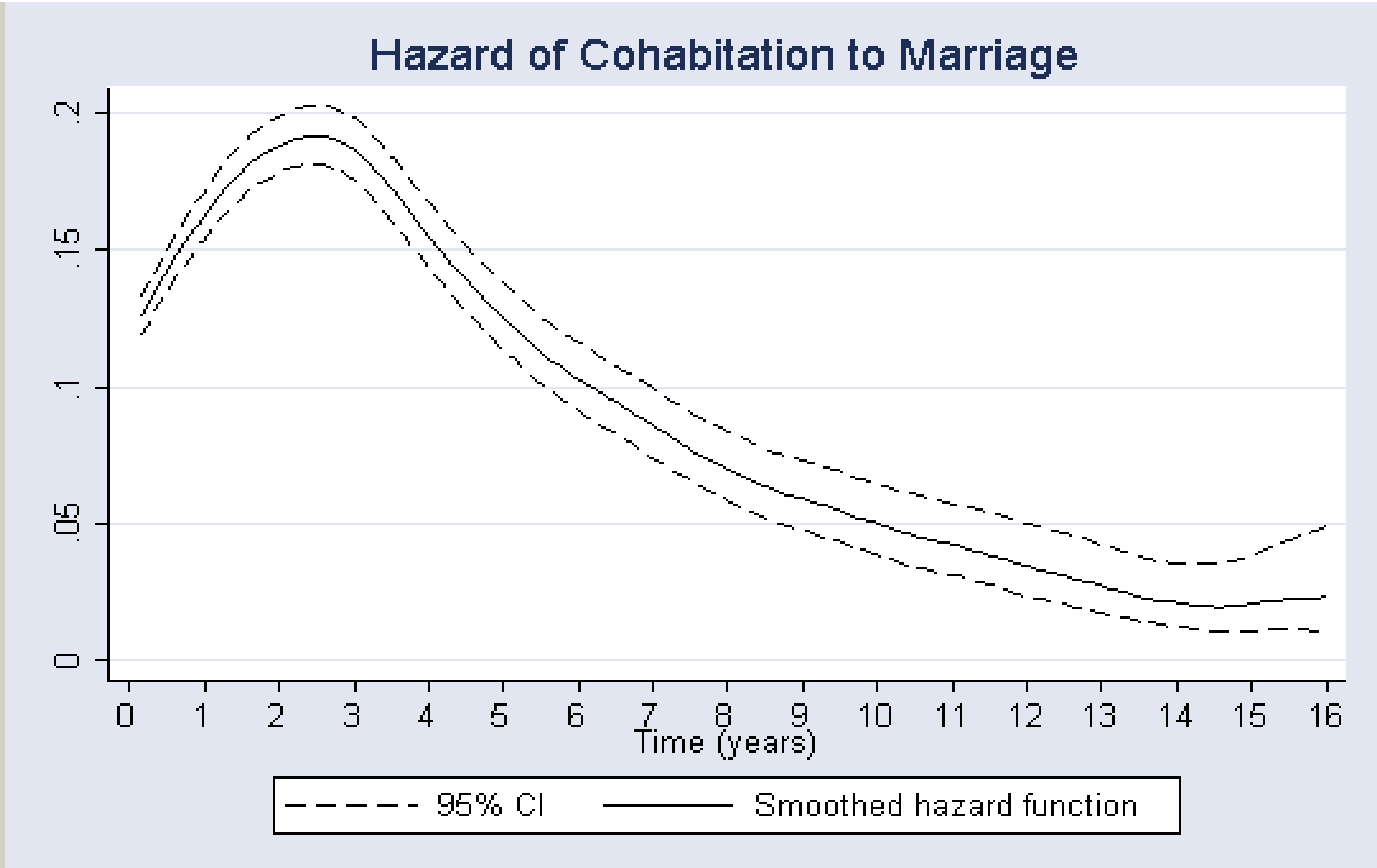
Could simulate values from $MVN\,(0, \hat{\Omega}_u)$

Competing Risks: Example

- Data from NCDS, women age 16-42
- Outcomes of cohabitation
  - Separation (r=1)
  - Marriage to cohabiting partner (r=2)
  - Staying cohabiting, i.e. “no event” (r=0)
- $\alpha^{(r)}(t)$ cubic polynomials



Hazard of Cohabitation to Separation
.02
.04
.06
.08
.1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Time (years)
95% CI
Smoothed hazard function


Hazard of Cohabitation to Marriage
0
.05
.1
.15
.2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Time (years)
95% CI
Smoothed hazard function

Years to Partnership Transitions: Quartiles

	25%	50%	75%
Cohab → Separation	3.5	9.1	-
Cohab → Marriage	1.3	2.9	10.3
Marriage → Separation	13.8	-	-

Cohabitation Outcomes: Selected Results

	Separation (r=1)		Marriage (r=2)	
	Est.	(SE)	Est.	(SE)
Age at start (ref=20-24)				
<20	0.09	(0.14)	-0.02	(0.09)
25-29	-0.42	(0.10)	-0.05	(0.06)
30-34	-0.41	(0.12)	-0.22	(0.08)
35+	-0.67	(0.16)	-0.35	(0.11)
Post-16 educ.				
1	0.21	(0.12)	-0.01	(0.08)
2	0.12	(0.14)	0.10	(0.08)
3-5	0.11	(0.13)	-0.04	(0.08)
6+	0.12	(0.15)	-0.11	(0.10)
$\mathrm{Var}(u_j^{(r)})$	0.59	(0.14)	0.23	(0.05)
$\mathrm{Cov}(u_j^{(1)}, u_j^{(2)})$	0.08 **Corr=0.22**	(0.06)		

Competing Risks and Multiple States: Example

- Jointly model outcomes of marital and cohabiting partnerships:
  - Marriage $\rightarrow$ separation
  - Cohabitation $\rightarrow$ separation
  - Cohabitation $\rightarrow$ marriage
- Multiple states: marriage and cohabitation
- Competing risks: 2 outcomes of cohabitation
- Model can be extended to consider *all* possible partnership transitions. Here we omit ***partnership formation*** (single $\rightarrow$ marriage, single $\rightarrow$ cohabitation)

Random Effects Covariance Matrix

	Mar→ Sep	Cohab→ Sep	Cohab→ Mar
Mar→ Sep	1.15*		
Cohab→Sep	0.46* corr=0.53	0.65*	
Cohab→Mar	0.12	0.08	0.28*

*95% interval estimate does not contain zero

Interpretation of Random Effect Correlation

Correlation between random effects for dissolution of marriage and cohabitation is estimated as 0.53.

Women with a high (low) risk of separation from marriage tend also to have a high (low) risk of separation from cohabitation.

Competing Risks and Multiple States: Another Example

(Steele et al. 2004, *Statistical Modelling*)

- Contraceptive use dynamics in Indonesia. Define episode of use as a continuous period of using the ***same*** method of contraception.
  - 2 states: use and nonuse
  - An episode of use can end in 2 ways: a transition to nonuse (discontinuation), or a switch to another method (a transition within the same state).

- Estimate 3 equations jointly: binary logit model for transition from nonuse to use, and multinomial logit model for transitions from use

Selected Results: Fixed Effects

	Use→nonuse Est. (SE)	Use →other method Est. (SE)	Nonuse →use Est. (SE)
Urban (ref.=rural)	0.13 (0.04)	0.06 (0.05)	0.26 (0.04)
SES (ref.=low)			
Medium	-0.12 (0.05)	0.35 (0.07)	0.24 (0.05)
High	-0.20 (0.05)	0.29 (0.08)	0.45 (0.05)

Random Effect Correlations from Two Different Models

	Use→nonuse	Use →other method	Nonuse →use
Use→nonuse	1		
Use →other method	0.020 0.011	1	
Nonuse →use	-0.783* -0.052	0.165* 0.095	1

Model 1: Duration effects only

Model 2: Duration + covariate effects

*Correlation significantly different from zero at 5% level

Random Effect Correlations: Interpretation

- In “duration effects only” model, there is a large negative correlation between random effects for nonuse $\rightarrow$ use and use $\rightarrow$ nonuse
  - Long durations of use associated with short durations of nonuse
- This is due to short episodes of postnatal nonuse followed by long episodes of use (to space or limit future births)
  - Correlation is effectively zero when we control for whether episode of nonuse follows a live birth (one of the covariates)

Independence of Irrelevant Alternatives (IIA)

- One potential problem of the multinomial logit model is the IIA assumption.

- The IIA assumption is that the hazard of one event relative to “no event” is independent of the hazards of each of the other events relative to “no event”.

- This may be unreasonable if some events can be regarded as similar.

Example: Partnership Formation (Marriage vs. Cohabitation)

- Under IIA, assume the hazard of cohabitation vs. staying single is uncorrelated with the hazard of marriage vs. staying single.
- E.g. If there is something unobserved (not in X's) that made marriage unfeasible, we assume those who would have married distribute themselves between cohabitation and single in the same proportions as those who originally chose not to marry.
- But as marriage and cohabitation are similar in some respects, we might expect those who are precluded from marriage to be more likely to cohabit rather than remain single (Hill, Axinn and Thornton, 1993, *Sociological Methodology*).

IIA: Solutions

- The random effects multinomial logit offers some protection against the IIA assumption. The correlation between the cause-specific random effects allows for similarity between alternatives due to **time-invariant** individual characteristics. It does not control for unmeasured factors that vary across episodes (e.g. whether respondent or partner is already married).

- Another approach is to use a multinomial probit model (e.g. using aML). Or nested logit model (Hill et al. 1993).

Multiple Processes

Endogeneity (1)

Consider a 2-level random effects model for a Normal response:

$$y_{ij} = \boldsymbol{\beta}\mathbf{x}_{ij} + u_j + e_{ij}$$

One assumption of the model is that the predictors $\mathbf{x}_{ij}$ are uncorrelated with the residuals (u_j,e_{ij}), i.e. we assume that $\mathbf{x}_{ij}$ are **exogenous**.

This may too strong an assumption for some predictors, particularly “choice” variables.

Endogeneity (2)

E.g. suppose y_{ij} is birth weight of child i of woman j, and z_{ij} (an element of $\mathbf{x}_{ij}$) is number of antenatal visits during pregnancy.

Some of the factors that influence birth weight may also influence the uptake of antenatal care; these may be characteristics of the particular pregnancy (e.g. woman's health during pregnancy) or of the woman (health-related behaviour). Some of these may be unobserved.

i.e. y and z are to some extent jointly determined. z is said to be endogenous.

This will lead to correlation between z and u and/or e, and a biased estimate of the coefficient of z.

Simultaneous Equations Model

One approach to allow for endogeneity is to model y and z jointly.

Estimate two equations:

$$y_{ij} = \boldsymbol{\beta}^{(y)} \mathbf{x}_{ij}^{(y)} + \gamma z_{ij} + u_j^{(y)} + e_{ij}^{(y)}$$

$$z_{ij} = \boldsymbol{\beta}^{(z)} \mathbf{x}_{ij}^{(z)} + u_j^{(z)} + e_{ij}^{(z)}$$

where we allow residuals at the same level to be correlated across equations:

$$\operatorname{cov}(u_j^{(y)}, u_j^{(z)}) = \sigma_{uyz}; \quad \operatorname{cov}(e_{ij}^{(y)}, e_{ij}^{(z)}) = \sigma_{eyz}$$

Because of these correlations, the equations must be estimated simultaneously (or in 2-stages in single-level case).

Simultaneous Equations Model: Identification

In order to identify the model, some exclusion restrictions must be placed on the covariates and/or random effect covariances:

- If covariances at both levels are permitted to be nonzero $x_{ij}^{(z)}$ must contain at least one variable not contained in $x_{ij}^{(y)}$

- If one covariance is assumed to equal zero, covariate exclusions are not strictly necessary for identification.

Simultaneous Equations Model: Estimation

Treat (y_{ij}, z_{ij}) as a bivariate response, and estimate multilevel bivariate response model (e.g. in *MLwiN*, Chapter 14 in manual).

Model can be extended to handle categorical responses (e.g. bivariate probit model) or mixture of Normal and categorical responses.

MLwiN can handle mixture of binary and Normal, and binary and multinomial (if zero covariance assumed at lowest level). *aML* can handle other mixtures.

Simultaneous Equations Model for Correlated Event Histories (1)

Suppose y is a duration variable and z is either also a duration or the *outcome* of an event history.

E.g. (Lillard 1993; Lillard and Waite, 1993)

y_{ij} is duration of marriage i of person j

$z_{ij}(t)$ is number of children from marriage i at time t (outcome of birth history)

The unobserved individual characteristics that affect hazard of marital separation may be correlated with those that affect hazard of birth (or conception) during marriage.

Simultaneous Equations Model for Correlated Event Histories (2)

$h_{ij}^{P}(t)$ Hazard of marital separation at time t

$h_{ij}^{F}(t)$ Hazard of conception (leading to live birth) at time t

P – partnership process
F – fertility process

Model hazards jointly using simultaneous (multiprocess) event history model.

Discrete-time Model

$$\operatorname{logit}[h_{ij}^{P}(t)] = \alpha^{P}(t) + \beta^{P} x_{ij}^{P}(t) + \gamma z_{ij}(t) + u_{j}^{P}$$

$$\operatorname{logit}[h_{ij}^{F}(t)] = \alpha^{F}(t) + \beta^{F} x_{ij}^{F}(t) + u_{j}^{F}$$

where $\operatorname{cov}(u_{j}^{P}, u_{j}^{F}) = \sigma_{uPF}$

Discrete-time Model: Estimation

For each time interval t define a bivariate response $(y_{ij}^{P}(t),\ y_{ij}^{F}(t))$

$y_{ij}^{P}(t)$ indicates partnership transition (e.g. marital separation)
- could be binary or multinomial (competing risks)

$y_{ij}^{F}(t)$ indicates fertility transition (conception)

Fit multilevel bivariate response model to allow for correlation between random effects.

Example: Lillard (1993)

Effect of Children from Current Marriage on Log-Hazard of Marital Separation

No. children (ref.=none)	$\sigma_{uPF} = 0$	σ_{uPF} free
1	-0.56 (0.10)	-0.33 (0.11)
2+	**-0.01 (0.05)**	**0.27 (0.07)**

$$\mathrm{Co\hat{r}r}(u_j^P, u_j^F) = -0.75\ (0.20)$$

Multiprocess Models with Multiple States and Competing Risks

Suppose we consider outcomes of cohabiting partnerships together with marital separation. This leads to multiple states (marriage & cohabitation) and competing risks (separation & marriage).

Partnership transition indicator will be binary for marriage and multinomial for cohabitation. Include state dummies and their interactions with covariates as before.

Add binary fertility response, leading to bivariate data with mixture of binary and multinomial responses.

Example: Steele et al. (2005)

Estimate 5 hazards equations for following transitions:

- marriage $\rightarrow$ separation
- cohabitation $\rightarrow$ separation
- cohabitation $\rightarrow$ marriage
- conception (leading to live birth) in marriage
- conception in cohabitation

Each equation includes a woman-specific random effect.

Random effects correlated across equations.

Random Effects Covariance Matrix: Partnership Transitions

	Mar→ Sep	Cohab→Sep	Cohab→Mar
Mar→ Sep	1.16*		
Cohab→Sep	0.49* **r=0.52**	0.77*	
Cohab→Mar	0.13 **r=0.21**	0.11 **r=0.23**	0.32*

*95% interval estimate does not contain zero

Random Effects Covariance Matrix: Fertility

	Marriage	Cohabitation
Marriage	0.05*	
Cohabitation	-0.01 **r=-0.06**	0.22*

*95% interval estimate does not contain zero

Selected Random Effect Correlations Across Processes

- Separation from marriage and cohabiting conception
  r = 0.42* (*sig. at 5% level)

- Separation from cohabitation and cohabiting conception
  r = 0.32*

- Cohabitation to marriage and cohabiting conception
  r = 0.30*

Effect of Selected Fertility Variables on Log-odds of Separation from Cohabitation

Age/Father[a]	Single Process	Multiprocess
Preschool/Curr[b]		
1	**-0.24***	**-0.29***
2+	**-0.75***	**-0.88***
Older/Curr		
1	**-0.03**	**-0.06**
2+	**0.24**	**0.14**
Preschool/Prev	-0.33	-0.34
Older/Prev	-0.01	-0.02
Non-coresid	-0.02	-0.02
Corr($u^{PC(1)}$,u^{FC})	-	0.32*

*95% interval estimate does not contain zero
[a]Father is current or previous partner.
[b]Reference category for all vars is 0 children.

Other Examples of Potentially Correlated Event Histories

- Partnership formation and childbearing
- Partnership formation and education
- Partnership formation/outcomes and employment
- Childbearing and employment
- Housing and employment

Bibliography:
Modelling the Time to a Single Event

Allison, P.D. (1982) "Discrete-time methods for the analysis of event history data." In Sociological Methodology (Ed. S. Leinhardt). San Francisco: Jossey-Bass.

Singer, J.D. and Willet, J.B. (1993) "It's about time: Using discrete-time survival analysis to study duration and the timing of events." *Journal of Educational Statistics*, **18**: 155-195.

Bibliography:

Models for Recurrent Events and Unobserved Heterogeneity

Davis, R.B., Elias, P. and Penn, R. (1992) “The relationship between a husband’s unemployment and a wife’s participation in the labour-force.” *Oxford Bulletin of Economics and Statistics*, **54**:145-71.

Kravdal, Ø. (2001) “The high fertility of college educated women in Norway: An artefact of the separate modelling of each parity transition.” *Demographic Research*, **5**, Article 6.

Bibliography:

Modelling Transitions Between Multiple States

Enberg, J., Gottschalk, P. and Wolf, D. (1990) “A random-effects logit model of work-welfare transitions.” *Journal of Econometrics*, **43**: 63-75.

Goldstein, H., Pan, H. and Bynner, J. (2004) “A flexible procedure for analysing longitudinal event histories using a multilevel model.” *Understanding Statistics*, **3**: 85-99.

Steele, F., Goldstein, H. and Browne, W. (2004) “A general multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics.” *Journal of Statistical Modelling*, **4**: 145-159.

Steele, F., Kallis, C., Goldstein, H. and Joshi, H. (2005) “The relationship between childbearing and transitions from marriage and cohabitation in Britain.” *Demography*, **42**: 647-673.

Bibliography:
Modelling Competing Risks

Enberg, J., Gottschalk, P. and Wolf, D. (1990) “A random-effects logit model of work-welfare transitions.” *Journal of Econometrics*, **43**: 63-75.

Hill, D.H., Axinn, W.G. and Thornton, A. (1993) “Competing hazards with shared unmeasured risk factors.” *Sociological Methodology*, **23**: 245-277.

Steele, F., Diamond, I. and Wang, D. (1996) “The determinants of the duration of contraceptive use in China: A multilevel multinomial discrete hazards modelling approach.” *Demography*, **33**: 12-33.

Steele, F., Goldstein, H. and Browne, W. (2004) “A general multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics.” *Journal of Statistical Modelling*, **4**: 145-159.

Bibliography:
Multiprocess Models for Correlated Histories

Brien, M.J., Lillard, L.A. and Waite, L.J. (1999) “Interrelated family-building behaviors: Cohabitation, marriage, and nonmarital conception.” *Demography*, **36**: 535-551.

Steele, F., Kallis, C., Goldstein, H. and Joshi, H. (2005) “The relationship between childbearing and transitions from marriage and cohabitation in Britain.” *Demography*, **42**: 647-673.

Upchurch, D.M., Lillard, L.A. and Panis, C.W.A. (2002) “Nonmarital childbearing: Influences of education, marriage, and fertility.” *Demography*, **39**: 311-329.

Waite, L.J. and Lillard, L.A. 1991. “Children and marital disruption.” *American Journal of Sociology*, **96**: 930-53.