Discrete-time Event History Analysis

LECTURES

Fiona Steele and Elizabeth Washbrook

Centre for Multilevel Modelling

University of Bristol

16-17 July 2013

Course outline

- Day 1:
  1. Introduction to discrete-time models: Analysis of the time to a single event
  2. Multilevel models for recurrent events and unobserved heterogeneity
- Day 2:
  3. Modelling transitions between multiple states
  4. Competing risks
  5. Multiprocess models

1. Analysis of time to a single event

What is event history analysis?

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

Event history analysis also known as:

- Survival analysis (especially in biostatistics and when events are not repeatable)
- Duration analysis
- Hazard modelling

Examples of applications

- Health. Age at death; duration of hospital stay
- Demography. Time to first birth (from when?); time to first marriage; time to divorce; time living in same house or area
- Economics. Duration of an episode of employment or unemployment
- Education. Time to leaving full-time education (from end of compulsory schooling); time to exit from teaching profession

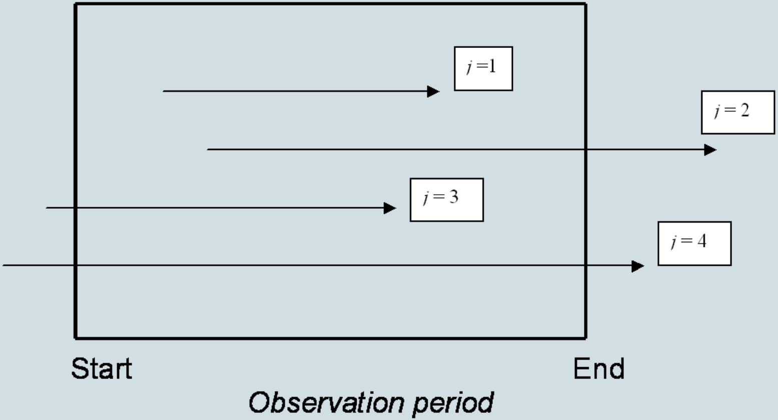
Types of event history data

- Dates of start of exposure period and events, e.g. dates of start and end of an employment spell
  - Usually collected retrospectively
  - Sources include panel and cohort studies (partnership, birth, employment and housing histories)
- Current status data from panel study, e.g. current employment status at each year
  - Collected prospectively

Special features of event history data

- Durations are always positive and their distribution is often positively skewed (long tail to the right)
- Censoring. There are usually people who have not yet experienced the event when we observe them, but may do so at an unknown time in the future
- Time-varying covariates. The values of some covariates may change over time

Types of censoring



Types of censoring

Line starts when individual becomes at risk of event.

Arrowhead indicates time that event occurs.

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

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

Right-censoring

- Right-censoring is the most common form of censoring. Durations are right-censored if the event has not occurred by the end of the observation period.
  - E.g. in a study of divorce, most respondents will still be married when last observed
- Excluding right-censored observations (e.g. still married) leads to bias and may drastically reduce sample size
- Usually assume censoring is non-informative

Right-censoring: Non-informative assumption

We retain right-censored observations under 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 (or dropped out) 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_i	event time for individual i
δ_i	censoring/event indicator $= 1$ if uncensored (i.e. observed to have event) $= 0$ if censored

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

Our outcome variable is $y_i = min(t_i, c_i)$.

Our observed data are (y_i, δ_i).

Descriptive Analysis

The hazard function

A key quantity in EHA is the hazard function:

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

where the numerator is the probability that an event occurs during a very small interval of time $[t, t + \Delta t)$, given that no event occurred before time t.

We divide by the width of the interval, Δt, to get a rate.

$h(t)$ is also known as the transition rate, the instantaneous risk, or the failure rate.

The survivor function

Another useful quantity in EHA is the survivor function:

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

the probability that an individual does not have the event before t, or ‘survives’ until at least t.

Its complement is the cumulative distribution function:

$$F(t) = 1 - S(t) = Pr(T < t)$$

the probability that an individual has the event before t.

Non-parametric estimation of $h(t)$

Group time so that t is now an interval of time (duration may already by grouped, e.g. in months or years).

$r(t)$ is number at 'risk' of experiencing event at start of interval t
$d(t)$ is number of events ('deaths') observed during t
$w(t)$ is number of censored cases ('withdrawals') in interval t

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

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

Note. Assumes censoring times are spread uniformly across interval t. Some estimators have $r(t) - 0.5w(t)$ as the denominator, or ignore censored cases.

Estimation of $S(t)$

The survivor function for interval t can be estimated from $\hat{h}(t)$ as:

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

E.g. probability of surviving to the start of 3rd interval

$=$ probability no event in 1st interval **and** no event in 2nd interval

$= \hat{S}(3) = [1-\hat{h}(1)] \times [1-\hat{h}(2)]$

Example: Time to 1st partnership

t	$r(t)$	$d(t)$	$w(t)$	$h(t)$	$S(t)$
16	500	9	0	0.02	1
17	491	20	0	0.04	0.98
18	471	32	0	0.07	0.94
19	439	52	0	0.12	0.88
20	387	49	0	0.13	0.77
.	.	.	.	.	.
.	.	.	.	.	.
32	39	3	0	0.08	0.08
33	36	1	35	0.03	0.07

Source: National Child Development Study (1958 birth cohort). Note that respondents were interviewed at age 33, so there is no censoring before then.

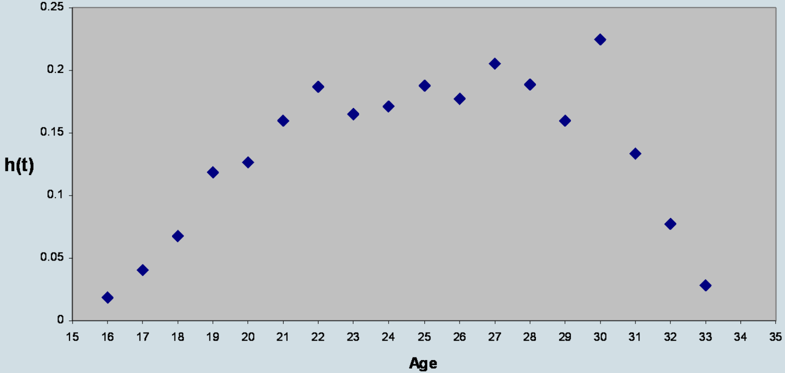
Example of interpretation

Event is partnering for the first time.

'Survival' here is remaining single.

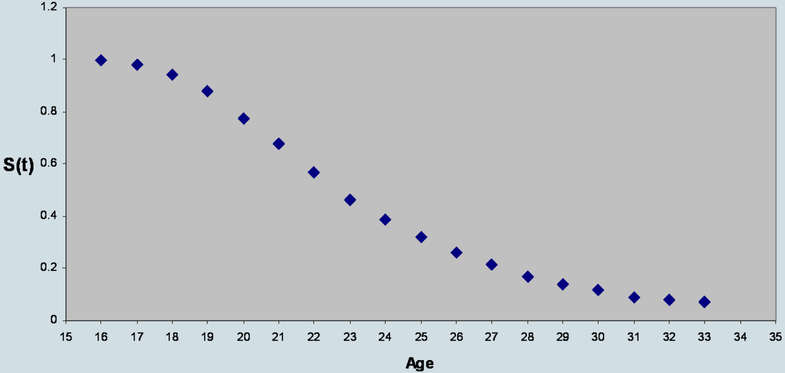
- $h(16) = 0.02$ so 2% partnered before age 17
- $h(20) = 0.13$ so, of those who were unpartnered at their 20th birthday, 13% partnered before age 21
- $S(20) = 0.77$ so 77% had not partnered by age 20

Hazard of 1st partnership



If an individual has not partnered by their late 20s, their chance of partnering declines thereafter.

Survivor function: Probability of remaining unpartnered



Note that the survivor function will always decrease with time.
The hazard function may go up and down.

Continuous-time Models

Introducing covariates: Event history modelling

There are many different types of event history model, which vary according to:

- Assumptions about the shape of the hazard function
- Whether time is treated as continuous or discrete
- Whether the effects of covariates can be assumed constant over time (proportional hazards)

The Cox proportional hazards model

The most commonly applied model is the Cox model which:

- Makes no assumptions about the shape of the hazard function
- Treats time as continuous
- Assumes that the effects of covariates are constant over time (although this can be modified)

The Cox proportional hazards model

$h_i(t)$ is the hazard for individual i at time t

$\mathbf{x}_i$ is a vector of covariates (for now assumed fixed over time) with coefficients β

$h_0(t)$ is the baseline hazard, i.e. the hazard when $\mathbf{x}_i = 0$

The Cox model can be written:

$$h_i(t) = h_0(t)\exp(\beta\mathbf{x}_i)$$

or sometimes as:

$$\log h_i(t) = \log h_0(t) + \beta\mathbf{x}_i$$

An individual's hazard depends on t through $h_0(t)$ which is left unspecified, so no need to make assumptions about the shape of the hazard.

Cox model: Interpretation (1)

$$h_i(t) = h_0(t)\exp(\beta \mathbf{x}_i)$$

Covariates have a multiplicative effect on the hazard.

For each 1-unit increase in x the hazard is multiplied by $\exp(\beta)$.

To see this, consider a binary x coded 0 and 1:

$$x_i = 0 \implies h_i(t) = h_0(t)$$

$$x_i = 1 \implies h_i(t) = h_0(t)\exp(\beta)$$

So $\exp(\beta)$ is the ratio of the hazard for $x = 1$ to the hazard for $x = 0$, called the relative risk or hazard ratio.

Cox model: Interpretation (2)

- $\exp(\beta) = 1$ implies no effect of x on the hazard
- $\exp(\beta) > 1$ implies a positive effect on the hazard, i.e. higher values of x are associated with shorter durations
  - e.g. $\exp(\beta) = 2.5$ implies an increase in $h(t)$ by a factor of $(2.5 - 1) \times 100 = 150\%$ for a 1-unit increase in x
- $\exp(\beta) < 1$ implies a negative effect on the hazard, i.e. lower values of x are associated with longer durations
  - e.g. $\exp(\beta) = 0.6$ implies a decrease in $h(t)$ by a factor of $(1 - 0.6) \times 100 = 40\%$ for a 1-unit increase in x

Example: Gender effects on age at 1st partnership

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	.400507	.0933196	4.29	0.000	.217604	.5834101

- The log-hazard of forming the 1st partnership at age t is 0.4 points higher for women than for men
- The hazard of forming the 1st partnership at age t is $\exp(0.40) = 1.49$ times higher for women than for men
- Women partner at a younger age than men

The proportional hazards assumption

Consider a model with a single covariate x and two individuals with different values denoted by x_1 and x_2.

The proportional hazards model is written:

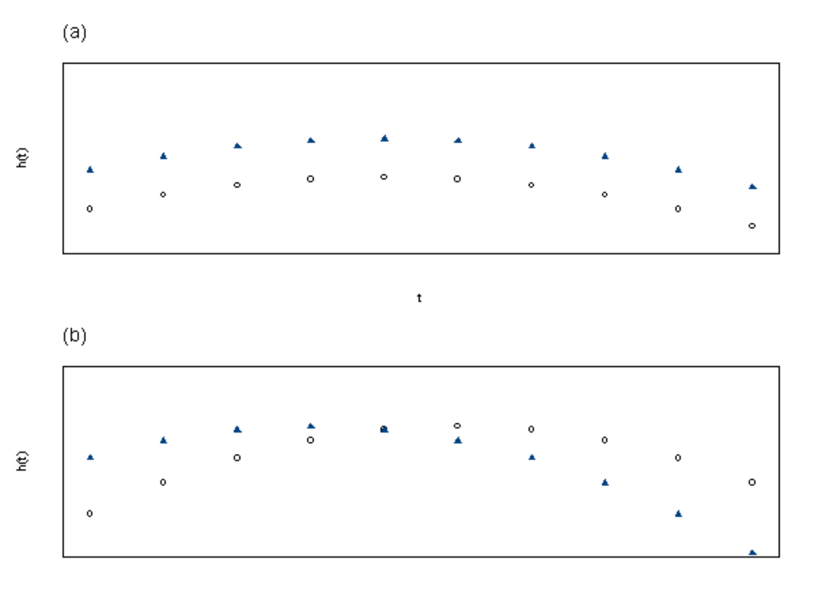
$$h_i(t) = h_0(t)\exp(\beta x_i)$$

So the ratio of the hazards for individual 1 to individual 2 is:

$$\frac{h_1(t)}{h_2(t)} = \frac{\exp(\beta x_1)}{\exp(\beta x_2)}$$

which does not depend on t. i.e. the effect of x is the same at all durations t.

Example of (a) proportional and (b) non-proportional hazards for binary x



Estimation of the Cox model

All statistical software packages have in-built procedures for estimating the Cox model. The input data are each individual's duration y_i and censoring indicator δ_i.

The data are restructured before estimation (although this is hidden from the user), and the Cox model is then estimated using Poisson regression.

We will look at this data restructuring to better understand the model and its relationship with the discrete-time approach. But note that you do not have to do this restructuring yourself!

Creation of risks sets

A risk set is defined for each observed event time and contains all individuals at risk of the event at that time.

Suppose there are K distinct uncensored event times and denote the ordered times by $t_{(1)}, t_{(2)}, \ldots, t_{(K)}$.

Example. Suppose ordered uncensored event times (age at marriage) are:

k	1	2	3	4	5	6
$t_{(k)}$	16	17	18	21	22	24

The event time ranges from 16 to 24, so there are potentially 9 event times (taking 16 as the origin). But there are 6 risk sets because no events were observed at $t = 19, 20, 23$.

Risk set based file

Consider records for 3 individuals:

individual i	y_i	δ_i
1	21	1
2	18	0
3	16	1

$$\downarrow$$

individual i	risk set k	$t_{(k)}$	y_{ki} (event at $t_{(k)}$)
1	1	16	0
1	2	17	0
1	3	18	0
1	4	21	1
2	1	16	0
2	2	17	0
2	3	18	0
3	1	16	1

Results from fitting Cox model

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	.393816	.0934761	4.21	0.000	.2106062	.5770258
fulltime	-1.031132	.1902864	-5.42	0.000	-1.404086	-.6581774

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	1.482628	.1385903	4.21	0.000	1.234426	1.780734
fulltime	.3566031	.0678567	-5.42	0.000	.2455913	.5177942

- Hazard of partnering at age t is $(1.48 - 1) \times 100 = 48\%$ higher for women than for men (i.e. W partner quicker than M)
- Being in full-time education decreases the hazard by $(1 - 0.36) \times 100 = 64\%$

Discrete-time Models

Discrete-time data

In social research, event history data are usually collected:

- retrospectively in a cross-sectional survey, where dates are recorded to the nearest month or year, *OR*
- prospectively in waves of a panel study (e.g. annually)

Both give rise to discretely-measured durations.

Also called interval-censored because we only know that an event occurred at some point during an interval of time.

Data preparation for a discrete-time analysis

We must first restructure the data.

We expand the event times and censoring indicator (y_i, δ_i) to a sequence of binary responses $\{y_{ti}\}$ where y_{ti} indicates whether an event has occurred in time interval $[t, t+1)$.

The required structure is very similar to the risk set-based file for the Cox model, but the user has to do the restructuring rather than the software.

Also we now have a record for *every* time interval (not risk sets, i.e. intervals where events occur).

Data structure: The person-period file

individual i	y_i	δ_i
1	21	1
2	33	0

$\downarrow$

individual i	t	y_{ti}
1	16	0
1	17	0
.	.	.
1	20	0
1	21	1
2	16	0
2	17	0
.	.	.
2	32	0
2	33	0

Discrete-time hazard function

Denote by p_{ti} the probability that individual i has an event during interval t, given that no event has occurred before the start of t.

$$p_{ti} = Pr(y_{ti} = 1 | y_{t-1,i} = 0)$$

p_{ti} is a discrete-time approximation to the continuous-time hazard function $h_i(t)$.

Call p_{ti} the discrete-time hazard function.

Discrete-time logit model

After expanding the data fit a binary response model to y_{ti}, e.g. a logit model:

$$\log\left(\frac{p_{ti}}{1-p_{ti}}\right) = \alpha \mathbf{D}_{ti} + \beta \mathbf{x}_{ti}$$

p_{ti} is the probability of an event during interval t

$\mathbf{D}_{ti}$ is a vector of functions of the cumulative duration by interval t with coefficients α

$\mathbf{x}_{ti}$ is a vector of covariates (time-varying or constant over time) with coefficients β

Modelling the time-dependency of the hazard

Changes in p_{ti} with t are captured in the model by $\alpha\mathbf{D}_{ti}$, the baseline hazard function.

$\mathbf{D}_{ti}$ has to be specified by the user. Options include:

Polynomial of order p

$$\alpha\mathbf{D}_{ti} = \alpha_0 + \alpha_1 t + \ldots + \alpha_p t^p$$

Step function

$$\alpha\mathbf{D}_{ti} = \alpha_1 D_1 + \alpha_2 D_2 + \ldots + \alpha_q D_q$$

where $D_1, \ldots, D_q$ are dummies for time intervals $t = 1, \ldots, q$ and q is the maximum observed event time. If q large, categories may be grouped to give a piecewise constant hazard model.

Discrete-time analysis of age at 1st partnership

Two covariates: FEMALE and FULLTIME (time-varying)

We consider two forms of $\alpha \mathbf{D}_{ti}$:

- Step function: dummy variable for each year of age, 16-33
- Quadratic function: include t and t^2 as explanatory variables

Duration effects fitted as a step function

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
t2	.1343721	.4328896	0.31	0.756	-.7140759	.9828201
t3	.7052694	.4093336	1.72	0.085	-.0970098	1.507549
t4	1.097346	.406424	2.70	0.007	.3007698	1.893923
t5	1.209112	.4070036	2.97	0.003	.4113999	2.006825
t6	1.516955	.4037025	3.76	0.000	.7257128	2.308197
t7	1.440981	.4190112	3.44	0.001	.6197341	2.262228
t8	1.308467	.4290047	3.05	0.002	.4676332	2.149301
t9	1.366649	.4350251	3.14	0.002	.5140154	2.219282
t10	1.477445	.4401949	3.36	0.001	.6146787	2.340211
t11	1.398869	.4534112	3.09	0.002	.5101994	2.287539
t12	1.596732	.4584545	3.48	0.000	.6981773	2.495286
t13	1.498472	.4797737	3.12	0.002	.5581329	2.438811
t14	1.279797	.5111952	2.50	0.012	.2778724	2.281721
t15	1.704258	.5025129	3.39	0.001	.7193507	2.689165
t16	1.072423	.5881356	1.82	0.068	-.0803015	2.225147
t17	.4545552	.7175246	0.63	0.526	-.9517672	1.860878
t18	-.6196484	1.087373	-0.57	0.569	-2.750861	1.511564
female	.4678989	.1022771	4.57	0.000	.2674395	.6683582
fulltime	-1.132822	.1968646	-5.75	0.000	-1.51867	-.746975
_cons	-3.129274	.395101	-7.92	0.000	-3.903658	-2.35489

Reference category for *t* is age 16 (could have fitted dummies for all ages, t1-t18, and omitted intercept)

Comparison of Cox and logit estimates for age at 1st partnership

	Cox		Logit	
Variable	$\hat{\beta}$	se($\hat{\beta}$)	$\hat{\beta}$	se($\hat{\beta}$)
Female	0.394	0.093	0.468	0.102
Fulltime(t)	−1.031	0.190	−1.133	0.197

Same substantive conclusions, but:

- Cox estimates are effects on log scale, and exp(β) are hazards ratios (relative risks)
- Logit estimates are effects on log-odds scale, and exp(β) are hazard-odds ratios

When will Cox and logit estimates be similar?

- In general, Cox and logit estimates will get closer as the hazard function becomes smaller because:

  $\log(h(t)) \approx \log\left(\frac{h(t)}{1-h(t)}\right)$ as $h(t) \to 0$.

  The discrete-time hazard will get smaller as the width of the time intervals become smaller.

- A discrete-time model with a complementary log-log link, $\log(-\log(1-p_t))$, is an approximation to the Cox proportional hazards model, and the coefficients are directly comparable.

Duration effects fitted as a quadratic

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
t	.3671895	.0558752	6.57	0.000	.257676	.476703
tsq	-.0184918	.00315	-5.87	0.000	-.0246657	-.0123179
female	.468567	.102215	4.58	0.000	.2682292	.6689048
fulltime	-1.128162	.1861813	-6.06	0.000	-1.493071	-.7632535
_cons	-3.368389	.2371865	-14.20	0.000	-3.833266	-2.903512

Approximating step function by a quadratic leads to little change in estimated covariate effects.

Estimates from step function model were 0.468 (SE $=$ 0.102) for Female and -1.133 (SE $=$ 0.197) for Fulltime.

Non-proportional hazards

- So far we have assumed that the effects of x are the same for all values of t
- It is straightforward to relax this assumption in a discrete-time model by including interactions between x and t in the model
- Test for non-proportionality by testing the null hypothesis that the coefficients of the interactions between x and t are all equal to zero, using likelihood ratio test

Allowing and testing for non-proportional effects of gender on timing of 1st partnership

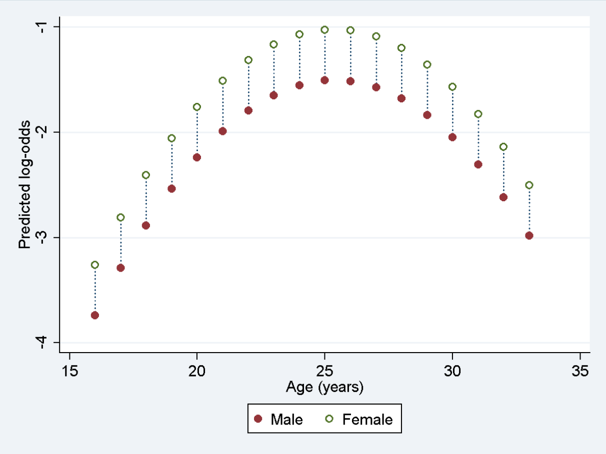
Add interactions: t_fem $= t \times$ female and tsq_fem $= t^2 \times$ female

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
t	.4283188	.0808491	5.30	0.000	.2698574	.5867801
tsq	-.0196737	.0043619	-4.51	0.000	-.0282229	-.0111245
female	1.022466	.4104085	2.49	0.013	.2180805	1.826852
t_fem	-.0651378	.1066576	-0.61	0.541	-.2741829	.1439072
tsq_fem	-.0011804	.0062517	-0.19	0.850	-.0134334	.0110726
fulltime	-1.114365	.1862045	-5.98	0.000	-1.479319	-.7494105
_cons	-3.764402	.3436495	-10.95	0.000	-4.437942	-3.090861

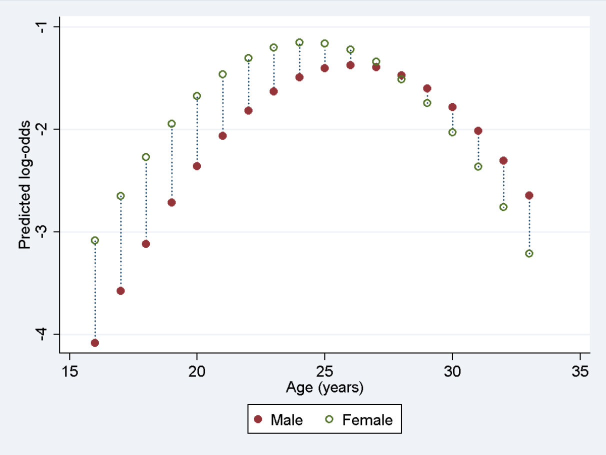
Likelihood-ratio statistic comparing main effects and interaction models $= 2(1332.5 - 1328.0) = 9$ on 2 d.f., p-value $= 0.011$

Conclude that interaction effects are significantly different from zero, i.e. effect of gender is non-proportional

Predicted log-odds of partnering: Proportional gender effects



Predicted log-odds of partnering: Non-proportional gender effects



2. Multilevel models for recurrent events and unobserved heterogeneity

Unobserved Heterogeneity

What is unobserved heterogeneity?

- Some individuals will be at higher risk of an event than others, and it is unlikely the reasons for this variability will be fully captured by covariates
- The presence of unmeasured individual-specific (time-invariant) risk factors leads to unobserved heterogeneity in the hazard
- Unobserved heterogeneity is also referred to as frailty, especially in biostatistics (more ‘frail’ individuals have a higher mortality risk)

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 the individual-level hazards.

For example, even if the hazards of individuals in a population are constant over time, the population hazard (averaged across individuals) will be time-dependent, typically decreasing.

This may be explained by a selection effect operating on individuals.

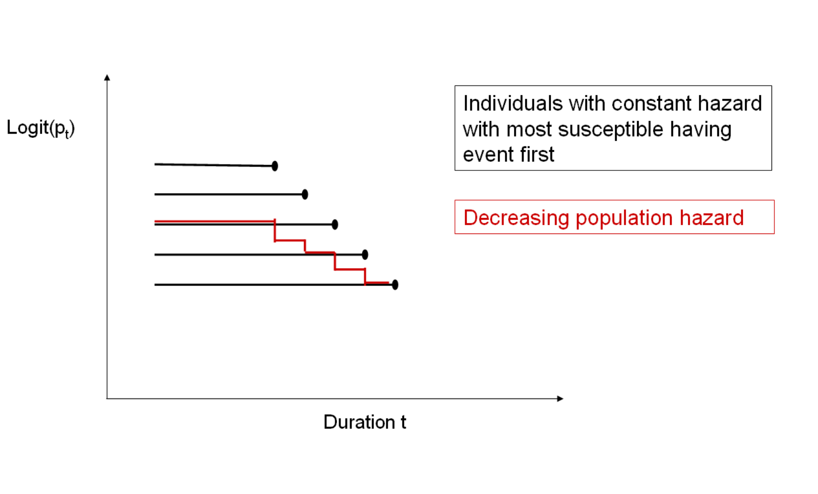
Selection effect of unobserved heterogeneity

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

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).

Illustration of selection for constant individual hazards



Impact of unobserved heterogeneity on duration effects

If unobserved heterogeneity is incorrectly ignored:

- A positive duration dependence will be understated (so an increasing baseline hazard will increase more sharply after accounting for UH)
- A negative duration dependence will be overstated

Note also that coefficients from random effects and traditional logit models 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:

$$\log\left(\frac{p_{ti}}{1-p_{ti}}\right) = \alpha\mathbf{D}_{ti} + \beta\mathbf{x}_{ti} + u_i$$

p_{ti} is the probability of an event during interval t

$\mathbf{D}_{ti}$ is a vector of functions of the cumulative duration by interval t with coefficients α

$\mathbf{x}_{ti}$ a vector of covariates with coefficients β

$u_i \sim N(0, \sigma_u^2)$ allows for unobserved heterogeneity ('frailty') between individuals due to time-invariant omitted variables

Estimation of discrete-time model with unobserved heterogeneity

- We can view the person-period dataset as a 2-level structure with time intervals (t) nested within individuals (i)
- The discrete-time logit model with a random effect u_i to capture unobserved heterogeneity between individuals is an example of a 2-level random intercept logit model
- The model can be fitted using routines/software for multilevel binary outcomes, e.g. Stata `xtlogit`

Results from analysis of 1st partnership without (1) and with (2) unobserved heterogeneity

	Model 1		Model 2	
	Est.	(SE)	Est.	(SE)
t	0.367	(0.056)	0.494	(0.122)
t^2	−0.018	(0.003)	−0.020	(0.004)
Female	0.469	(0.102)	0.726	(0.215)
Fulltime	−1.128	(0.186)	−1.187	(0.208)
Cons	−3.368	(0.237)	−4.134	(0.646)
σ_u	–	–	0.920	(0.400)

Likelihood-ratio test statistic for test of $H_0 : \sigma_u = 0$ is 3.74 on 1 d.f., p=0.027 (one-sided test as σ_u must be non-negative).

More on comparing coefficients from random effects and single-level logit models

In our analysis of age at 1st partnership, we saw that the positive effect of age ('duration') was understated if unobserved heterogeneity was ignored (as in Model 1).

Note also, however, that the effects of Female and Fulltime have also changed. In both cases, the magnitude of the coefficients has increased after accounting for unobserved heterogeneity.

This can be explained by a scaling effect.

Scaling effect of introducing u_i (1)

To see the scaling effect, consider the latent variable (threshold) representation of the discrete-time logit model.

Consider a latent continuous variable y^* that underlies observed binary y such that:

$$y_{ti} = \begin{cases} 1 & \text{if } y^*_{ti} \geq 0 \\ 0 & \text{if } y^*_{ti} < 0 \end{cases}$$

Threshold model

$$y^*_{ti} = \alpha \mathbf{D}_{ti} + \beta \mathbf{x}_{ti} + u_i + e^*_{ti}$$

- $e^*_{ti} \sim$ standard logistic (with variance $\simeq 3.29$) $\rightarrow$ logit model
- $e^*_{ti} \sim N(0, 1) \rightarrow$ probit model

So the level 1 residual variance, $\text{var}(e^*_{ti})$, is fixed.

Scaling Effect of Introducing u_i (2)

Single-level logit model expressed as a threshold model:

$$y^*_{ti} = \alpha \mathbf{D}_{ti} + \beta \mathbf{x}_{ti} + e^*_{ti}$$

$$\text{var}(y^*_{ti}|x_{ti}) = \text{var}(e^*_{ti}) = 3.29$$

Now add random effects:

$$y^*_{ti} = \alpha \mathbf{D}_{ti} + \beta \mathbf{x}_{ti} + u_i + e^*_{ti}$$

$$\text{var}(y^*_{ti}|x_{ti}) = \text{var}(u_i) + \text{var}(e^*_{ti}) = \sigma^2_u + 3.29$$

Adding random effects has increased the residual variance
$\rightarrow$ scale of y^* stretched out
$\rightarrow$ α and β increase in absolute value.

Scaling Effect of Introducing u_i (3)

Denote by β^{RE} the coefficient from a random effects model, and β^{SL} the coefficient from the corresponding single-level model.

The approximate relationship between these coefficients (for a logit model) is:

$$\beta^{RE} = \beta^{SL}\sqrt{\frac{\sigma_u^2 + 3.29}{3.29}}$$

Replace 3.29 by 1 to get expression for relationship between probit coefficients.

Note that the same relationship would hold for duration effects α if there was no selection effect. In general, both selection and scaling effects will operate on α.

Time to 1st partnership: Interpretation of coefficients from the frailty model

- For a given individual the odds of entering a partnership at age t when in FT education are $\exp(-1.19) = 0.30$ times the odds when not in FT 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 $\exp(0.73) = 2.08$ times higher for a woman than for a man
  - This interpretation is less useful, but we can 'average out' random effect to obtain population-averaged predicted probabilities

Population-averaged predicted probabilities

The probability of an event in interval t for individual i is:

$$p_{ti} = \frac{\exp(\alpha \mathbf{D}_{ti} + \beta \mathbf{x}_{ti} + u_i)}{1 + \exp(\alpha \mathbf{D}_{ti} + \beta \mathbf{x}_{ti} + u_i)}$$

where we substitute estimates of α, β, and u_i to get predicted probabilities.

Rather than calculating probabilities for each record ti, however, we often want predictions for specific values of $\mathbf{x}$. We do this by 'averaging out' the individual unobservables u_i.

Population-averaged predictions via simulation

Suppose we have 2 covariates, x_1 and x_2, and we want the mean predicted p_t for values of x_1 holding x_2 constant.

To get predictions for $t = 1, \ldots, q$ and $x_1 = 0, 1$:

1. Set $t = 1$ and x_{1ti}=0 for each record ti, retaining observed x_{2ti}
2. Generate u_i for each individual i from $N(0, \hat{\sigma}_u^2)$
3. Compute predicted p_t for each record ti based on $x_{1ti} = 0$, observed x_{2ti}, generated u_i, and $(\hat{\alpha}, \hat{\beta})$
4. Take mean of predictions to get mean p_t for $t = 1$ and $x_1 = 0$
5. Repeat 1-4 for $t = 2, \ldots, q$
6. Repeat 1-5 for $x_{1ti} = 1$

Recurrent Events

Multilevel event history data

Multilevel event history data arise when events are repeatable (e.g. births, partnership dissolution) or individuals are organised in groups.

Suppose events are repeatable, and define an **episode** as a continuous period for which an individual is at risk of experiencing an event, e.g.

Event	Episode duration
Birth	Duration between birth $k-1$ and birth k
Marital dissolution	Duration of marriage

Denote by y_{ij} the duration of episode i of individual j, which is fully observed if an event occurs ($\delta_{ij} = 1$) and right-censored if not ($\delta_{ij} = 0$).

Data structure: the person-period-episode file

individual j	episode i	y_{ij}	δ_{ij}
1	1	2	1
1	2	3	0

$$\downarrow$$

individual j	episode i	t	y_{tij}
1	1	1	0
1	1	2	1
1	2	1	0
1	2	2	0
1	2	3	0

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 **all** episodes, e.g. 'taste for stability' may influence risk of leaving a job.

The presence of such unobservables, and failure to account for them in the model, will lead to correlation between durations of episodes from the same individual.

Multilevel discrete-time model for recurrent events

Multilevel (random effects) discrete-time logit model:

$$\log\left(\frac{p_{tij}}{1-p_{tij}}\right) = \alpha \mathbf{D}_{tij} + \beta \mathbf{x}_{tij} + u_j$$

p_{tij} is the probability of an event during interval t

$\mathbf{D}_{tij}$ is a vector of functions of the cumulative duration by interval t with coefficients α

$\mathbf{x}_{tij}$ a vector of covariates (time-varying or defined at the episode or individual level) with coefficients β

$u_j \sim N(0, \sigma_u^2)$ allows for unobserved heterogeneity ('shared frailty') between individuals due to time-invariant omitted variables

Multilevel model for recurrent events: Notes

- The model for recurrent events is essentially the same as the (single-level) model for unobserved heterogeneity
  - Both can be estimated using multilevel modelling software/routines

- Recurrent events allow better identification of the random effect variance σ_u^2

- Allow for non-proportional effects of covariate x by including interaction between x and functions of t in $\mathbf{D}$

- Can allow duration and covariate effects to vary across episodes
  - Include a dummy for order of event and interact with t and x

Example: Women's employment transitions

- Analyse duration of non-employment (unemployed or out of labour market) episodes
  - Event is entry (1st episode) or re-entry (2nd + episodes) into employment
- Data are subsample from British Household Panel Study (BHPS): 1399 women and 2284 episodes
- Durations grouped into years $\Rightarrow$ 15,297 person-year records
- Baseline hazard is step function with yearly dummies for durations up to 9 years, then single dummy for 9+ years
- Covariates include time-varying indicators of number and age of children, age, marital status and characteristics of previous job (if any)

Multilevel logit results for transition to employment: Baseline hazard and unobserved heterogeneity

Variable	Est.	(se)
Duration non-employed (ref is < 1 year)		
[1,2) years	-0.646*	(0.104)
[2,3)	-0.934*	(0.135)
[3,4)	-1.233*	(0.168)
[4,5)	-1.099*	(0.184)
[5,6)	-0.944*	(0.195)
[6,7)	-1.011*	(0.215)
[7,8)	-1.238*	(0.249)
[8,9)	-1.339*	(0.274)
≥ 9 years	-1.785*	(0.175)
σ_u (SD of woman random effect)	0.662*	(0.090)

* $p < 0.5$

Multilevel logit results for transition to employment: Presence and age of children

Variable	Est.	(se)
Imminent birth (within 1 year)	−0.842*	(0.125)
No. children age $\leq$ 5 yrs (ref=0)		
1 child	−0.212*	(0.097)
$\geq$ 2	−0.346*	(0.143)
No. children age $>$ 5 yrs (ref=0)		
1 child	0.251	(0.118)
$\geq$ 2	0.446*	(0.117)

* $p < 0.5$

Multilevel logit analysis of employment: Main conclusions

- Unobserved heterogeneity. Significant variation between women. Deviance = 23.5 on 1 df; $p<0.01$
- Duration effects. Probability of getting a job decreases with duration out of employment
- Presence/age of children. Probability of entering employment lower for women who will give birth in next year or with young children, but higher for those with older children
- Other covariates. Little effect of age, but increased chance of entering employment for women who are cohabiting, have previously worked, whose last job was full-time, and whose occupation is 'professional, managerial or technical'

Grouping time intervals

When we move to more complex models, a potential problem with the discrete-time approach is that the person-period file can be very large (depending on sample size and length of the observation period relative to the width of discrete-time intervals).

It may be possible to group time intervals, e.g. using 6-month rather than monthly intervals.

BUT we must assume the hazard and values of covariates are constant within grouped intervals.

Analysing grouped intervals

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

e.g. for any 6-month interval some individuals will have the event or be censored after the 1st month while others will be exposed for the full 6 months.

Denote by n_{tij} the exposure time in grouped interval t of episode i for individual j . (**Note:** Intervals do not need to be the same width.)

Fit binomial logit model for grouped binary data, with response y_{tij} and denominator n_{tij} (e.g. using the `binomial()` option in the Stata `xtmelogit` command)

Example of grouped time intervals

Suppose an individual is observed to have an event during the 17th month of exposure, and we group durations into six-month intervals (t). Instead of 17 monthly records we would have three six-monthly records:

j	i	t	n_{tij}	y_{tij}
1	1	1	6	0
1	1	2	6	0
1	1	3	5	1

Software for Recurrent Events

- Essentially multilevel models for binary responses
- Mainstream software: e.g. Stata (`xtlogit`), SAS (`PROC NLMIXED`)
- Specialist multilevel modelling software: e.g. MLwiN (also via `runmlwin` in Stata), SABRE, aML

3. Modelling transitions between multiple states

States in Event Histories

In the models considered so far, there is a single event (or transition) 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 the analysis of transitions into employment we model the duration in the non-employment state
- In a study of marital dissolution we model the duration in the marriage state

More generally, we may wish to model transitions in the other direction (e.g. into non-employment or marriage formation) and possibly other transitions.

Examples of Multiple States

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

Examples:

- Employment states: employed full-time, employed part-time, unemployed, out of the labour market
- Partnership states: marriage, cohabitation, single (not in co-residential union)

We will begin with models for transitions between two states, e.g. non-employment (NE) $\leftrightarrow$ employment (E)

Transition Probabilities for Two States

Suppose there are two states indexed by s ($s = 1, 2$), and S_{tij} indicates the state occupied by individual j during interval t of episode i.

Denote by y_{tij} a binary variable indicating whether any transition has occurred during interval t, i.e. from state 1 to 2 or from state 2 to 1.

The probability of a transition from state s during interval t, given that no transition has occurred before the start of t is:

$$p_{stij} = Pr(y_{tij} = 1 | y_{t-1,ij} = 0, S_{tij} = s), \qquad s = 1, 2$$

Call p_{stij} a transition probability or discrete-time hazard for state s.

Event History Model for Transitions between 2 States

Multilevel two-state logit model:

$$\log\left(\frac{p_{stij}}{1-p_{stij}}\right) = \alpha_s \mathbf{D}_{stij} + \beta_s \mathbf{x}_{stij} + u_{sj},$$

p_{stij} is the probability of a transition from state s during interval t

$\mathbf{D}_{stij}$ is a vector of functions of cumulative duration in state s by interval t with coefficients α_s

$\mathbf{x}_{stij}$ a vector of covariates affecting the transition from state s with coefficients β_s

u_{sj} allows for unobserved heterogeneity between individuals in their probability of moving from state s. Assume $\mathbf{u}_j = (u_{1j}, u_{2j}) \sim$ bivariate normal.

Random Effect Covariance in a Two-State Model

We assume the state-specific random effects u_{sj} follow a bivariate normal distribution to allow for correlation between the unmeasured time-invariant influences on each transition.

For example, a highly employable person may have a low chance of leaving employment and a high chance of entering employment, leading to $\text{cov}(u_{1j}, u_{2j}) < 0$.

Allowing for $\text{cov}(u_{1j}, u_{2j}) \neq 0$ means that the equations for states $s = 1, 2$ must be estimated jointly. Estimating equations separately assumes that $\text{cov}(u_{1j}, u_{2j}) = 0$.

Data Structure for Two-State Model (1)

Start with an episode-based file.

E.g. employment (E) $\leftrightarrow$ non-employment (NE) transitions

j	i	State$_{ij}$	t_{ij}	δ_{ij}	Age$_{ij}$
1	1	E	3	1	16
1	2	NE	2	0	19

Note: (i) t in years; (ii) $\delta_{ij} = 1$ if a transition (event) occurs, 0 if censored; (iii) Age in years at start of episode

Data Structure for Two-State Model (2)

Convert episode-based file to discrete-time format with one record per interval t:

t	y_{tij}	E_{ij}	NE_{ij}	$E_{ij}Age_{ij}$	$NE_{ij}Age_{ij}$
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

Note: E_{ij} a dummy for employment, NE_{ij} a dummy for non-employment.

Example: Non-Employment $\leftrightarrow$ Employment

- corr(u_{1j}, u_{2j})=0.59, se=0.13, so large positive residual correlation between E $\rightarrow$ NE and NE $\rightarrow$ E
  - Women with high chance of *entering* E tend to have a high chance of *leaving* E
  - Women with low chance of *entering* E tend to have a low chance of *leaving* E
- Positive correlation arises from two sub-groups: short spells of E and NE, and longer spells of both types

Comparison of Selected Coefficients for NE $\rightarrow$ E

Only coefficients of covariates relating to employment history change:

	Single-state	Multistate
Ever worked	2.936	2.677
Previous job part-time	−0.441	−0.460

So positive effect of ‘ever worked’ has weakened, and negative effect of ‘part-time’ has strengthened.

Why Decrease in Effect of 'Ever Worked' on NE $\rightarrow$ E?

Direction of change from single-state to multistate (2.936 to 2.677) is in line with positive corr(u_{1j}, u_{2j}) in multistate model.

- Women in 'ever worked' must have made E $\rightarrow$ NE transition
- Positive correlation between E $\rightarrow$ NE and NE $\rightarrow$ E leads to disproportionate presence of women with high NE $\rightarrow$ E rate among 'ever worked'
- These women push up odds of NE $\rightarrow$ E among 'ever worked' (inflating estimate) if residual correlation uncontrolled

Why Increase in Effect of Previous Part-Time Job on NE $\rightarrow$ E?

Strengthening of negative effect when moving to the multistate model (-0.441 to -0.460) is also in line with positive $\text{corr}(u_{1j}, u_{2j})$.

- Women with tendency towards less stable employment (with high rate of E $\rightarrow$ NE) selected into part-time work
- Positive correlation between E $\rightarrow$ NE and NE $\rightarrow$ E leads to disproportionate presence of women with high NE $\rightarrow$ E rate in 'previous PT' category
- These women push up odds of NE $\rightarrow$ E in 'previous PT' (reducing 'true' negative effect of PT) if residual correlation uncontrolled

Autoregressive Models for Two States

An alternative way of modelling transitions between states is to include the lagged response as a predictor rather than the duration in the current state.

The response y_{tij} now indicates the state occupied at the start of interval t rather than whether a transition has occurred, i.e.

$$y_{tij} = \begin{cases} 1 & \text{if in state 1} \\ 0 & \text{if in state 2} \end{cases}$$

1st Order Autoregressive Model

An AR(1) model for the probability that individual j is in state 1 at t, p_{tj} is:

$$\log\left(\frac{p_{tj}}{1-p_{tj}}\right) = \alpha + \beta \mathbf{x}_{tj} + \gamma y_{t-1,j} + u_j$$

α is an intercept term

γ is the effect of the state occupied at $t-1$ on the log-odds of being in state 1 at t

$u_j \sim N(0, \sigma_u^2)$ is an individual-specific random effect

Interpretation of AR(1) Model

Suppose states are employment and unemployment. Common to find those who have been unemployed in the past are more likely to be unemployed in the future. Three potential explanations:

- A causal effect of unemployment at $t-1$ on being unemployed at t (state dependence γ)
- Unobserved heterogeneity, i.e. unmeasured individual characteristics affecting unemployment probability at all t (stable traits u_j)
- Non-stationarity, e.g. seasonality (not in current model)

The AR(1) model is commonly referred to as a state dependence model.

Transition Probabilities from the AR(1) Model

We model $p_{tj} = Pr(\text{state 1 at start of interval } t) = Pr(y_{tj} = 1)$

Suppose we fix $\mathbf{x}_{tj} = \mathbf{0}$ and $u_j = 0$.

Probability of moving from state 1 to 2

$$\begin{aligned} Pr(y_{tj} = 0 | y_{t-1,j} = 1) &= 1 - Pr(y_{tj} = 1 | y_{t-1,j} = 1) \\ &= 1 - \tfrac{\exp(\alpha+\gamma)}{1+\exp(\alpha+\gamma)} \end{aligned}$$

Probability of moving from state 2 to 1

$$Pr(y_{tj} = 1 | y_{t-1,j} = 0) = \frac{\exp(\alpha)}{1 + \exp(\alpha)}$$

Initial Conditions (1)

y may not be measured at the start of the process, e.g. we may not have entire employment histories.

Can view as a missing data problem. Suppose we observe y at the start of T intervals:

Observed	$(y_1, \ldots, y_T)$
Actual	$(y_{-k}, \ldots, y_0, y_1, \ldots, y_T)$

where first $k+1$ measures are missing.

We need to specify a model for y_1 (not just condition on y_1).

Initial Conditions (2)

In a random effects framework, we can specify a model for y_{1j} and estimate jointly with the model for $(y_{2j}, \ldots, y_{Tj})$, e.g.

$$\begin{aligned}\text{logit}(p_{1j}) &= \alpha_1 + \beta_1 \mathbf{x}_{t1j} + \lambda u_j \\ \text{logit}(p_{tj}) &= \alpha + \beta_1 \mathbf{x}_{tj} + \gamma y_{t-1,j} + u_j, \qquad t > 1\end{aligned}$$

Variants on the above are to set $\lambda = 1$ or to include different random effect in equation for $t = 1$, e.g. u_{1j}, and allow for correlation with u_j in equation for $t > 1$.

The inclusion of λ allows the between-individual residual variance to differ for $t = 1$ and $t > 1$.

Key Features of the AR(1) Model

- All relevant information about the process up to t is captured by y_{t-1} (1st order Markov assumption). This is why duration effects are not included.

- Because of the 1st order Markov assumption, there is no concept of an 'episode' (which is why we drop the i subscript)

- Effects of x (and time-invariant characteristics u_j) are the same for transitions from state 1 to 2, and from state 2 to 1

  - But it is straightforward to allow for transition-specific effects by interacting x with y_{t-1}

Which Model?

Consider AR(1) model when:

- Interested in separation of causal effect of y_{t-1} on y_t from unobserved heterogeneity
- Frequent movement between states (high transition probabilities)
- Duration in state at $t = 1$ is unknown, e.g. in panel data

Consider duration model when:

- Expect duration in state to have an effect on chance of transition
- More stable processes with long periods in the same state (low transition probabilities)

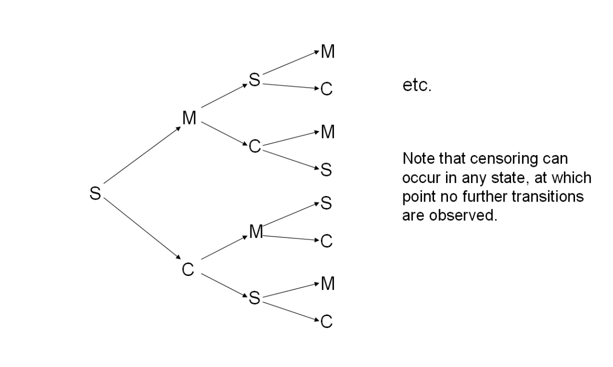
Software for Multiple States

- Two-state duration model can be framed as a multilevel random coefficients model
  - Coefficients for the two state dummies are intercepts which vary randomly across individuals to give random effects for each state
  - Software options as for one-state recurrent events model, but using `xtmelogit` in Stata
- Autoregressive model (with equation for $t = 1$) can also be fitted as a random coefficient model, but more general models (e.g. allowing different residual variances for $t = 1$ and $t > 1$) require specialist software such as Sabre and aML

4. Competing risks

More than Two states

In general there may be multiple states, possibly with different destinations from each state. E.g. consider transitions between marriage (M), cohabitation (C) and single (S).



Competing risks

We will begin with the special case where we are interested in transitions from one origin state, but there is more than one destination or type of transition.

Assume these destinations are mutually exclusive. We call these 'competing risks'.

Origin state	Competing risks
Alive	Different causes of death
Employed	Sacked, redundancy, switch job, leave labour market
Single	Marriage, cohabitation

Approaches to Modelling Competing Risks

Suppose there are R types of transition/event. For each interval t (of episode i of individual j) we can define a categorical response y_{tij}:

$$y_{tij} = \begin{cases} 0 & \text{if no event in } t \\ r & \text{if event of type } r \text{ in } t \ (r = 1, \ldots, R) \end{cases}$$

Analysis approaches

1. Multinomial model for y_{tij}
2. Define binary response $y_{tij}^{(r)}$ for event type r, treating all other types of event as censored. Analyse using multivariate response model

Multinomial Logit Model

Define $p_{tij}^{(r)} = \Pr(y_{tij} = r | y_{t-1,ij} = 0)$ for $r = 1, \ldots, R$.

Estimate R equations contrasting event type r with ‘no event’:

$$\log\left(\frac{p_{tij}^{(r)}}{p_{tij}^{(0)}}\right) = \alpha^{(r)}\mathbf{D}_{tij}^{(r)} + \beta^{(r)}\mathbf{x}_{tij}^{(r)} + u_j^{(r)}, \qquad r = 1, \ldots, R$$

where $(u_j^{(1)}, u_j^{(2)}, \ldots, u_j^{(R)}) \sim$ multivariate normal.

Correlated random effects allows for shared unobserved risk factors.

Multivariate Binary Response Model

In the second approach to modelling competing risks we define, for each interval t, R binary responses coded as:

$$y_{tij}^{(r)} = \begin{cases} 1 & \text{if event of type } r \text{ in } t \\ 0 & \text{if event of any type other than } r \text{ or no event in } t \end{cases}$$

and estimate equations for each event type:

$$\log\left(\frac{p_{tij}^{(r)}}{1-p_{tij}^{(r)}}\right) = \alpha^{(r)}\mathbf{D}_{tij}^{(r)} + \beta^{(r)}\mathbf{x}_{tij}^{(r)} + u_j^{(r)}, \qquad r = 1, \ldots, R$$

where $(u_j^{(1)}, u_j^{(2)}, \ldots, u_j^{(R)}) \sim$ multivariate normal.

Logic Behind Treating Other Events as Censored

Suppose we are interested in modelling partnership formation, where an episode in the 'single' state can end in marriage or cohabitation.

For each single episode we can think of durations to marriage and cohabitation, $t^{(M)}$ and $t^{(C)}$.

We cannot observe both of these. If a single episode ends in marriage, we observe only $t^{(M)}$ and the duration to cohabitation is censored at $t^{(M)}$. A person who marries is removed from the risk of cohabiting (unless they become single again).

For uncensored episodes we observe $\min(t^{(M)}, t^{(C)})$.

Comparing Methods

Coefficients and random effect variances and covariances will be different for the two models because the reference category is different:

- ‘No event’ in the multinomial model
  - Coefficients are effects on the log-odds of an event of type r relative to ‘no event’
- ‘No event + any event other than r’ in the multivariate binary model
  - Coefficients are effects on the log-odds of an event of type r relative to ‘no event of type r’

However, predicted transition probabilities will in general be similar for the two models.

Transition Probabilities

Multinomial model

$$p_{tij}^{(r)} = \frac{\exp(\alpha^{(r)}\mathbf{D}_{tij}^{(r)} + \beta^{(r)}\mathbf{x}_{tij}^{(r)} + u_j^{(r)})}{1 + \sum_{k=1}^{R} \exp(\alpha^{(k)}\mathbf{D}_{tij}^{(k)} + \beta^{(k)}\mathbf{x}_{tij}^{(k)} + u_j^{(k)})}$$

Multivariate binary model

$$p_{tij}^{(r)} = \frac{\exp(\alpha^{(r)}\mathbf{D}_{tij}^{(r)} + \beta^{(r)}\mathbf{x}_{tij}^{(r)} + u_j^{(r)})}{1 + exp(\alpha^{(r)}\mathbf{D}_{tij}^{(r)} + \beta^{(r)}\mathbf{x}_{tij}^{(r)} + u_j^{(r)})}$$

In each case, the 'no event' probability is $p^{(0)} = 1 - \sum_{k=1}^{R} p^{(k)}$.

To calculate probabilities for specific values of $\mathbf{x}$, substitute $u^{(r)} = 0$ or generate $u^{(r)}$ from multivariate normal distribution.

Example: Transitions to Full-time and Part-time Work

Selected results from bivariate model for binary responses, $y^{(FT)}$ and $y^{(PT)}$

	NE $\rightarrow$ FT		NE $\rightarrow$ PT	
Variable	Est	(se)	Est	(se)
Imminent birth	-1.19*	(0.18)	-0.26	(0.15)
1 kid ≤ 5	-1.27*	(0.15)	0.60*	(0.12)
2+ kids ≤ 5	-1.94*	(0.27)	0.81*	(0.17)
1 kid > 5	-0.42*	(0.19)	0.80*	(0.14)
2+ kids > 5	-0.26	(0.18)	1.24*	(0.15)

So having kids (especially young ones) reduces chance of returning to FT work, but increases chance of returning to PT work.

Example: Random Effect Covariance Matrix

	NE $\rightarrow$ FT	NE $\rightarrow$ PT
NE $\rightarrow$ FT	1.49 (0.13)	
NE $\rightarrow$ PT	-0.05 (0.11)	0.98 (0.11)

Note: Parameters on diagonal are standard deviations, and the off-diagonal parameter is the correlation. Standard errors in parentheses.

Correlation is not significant (deviance test statistic is < 1 on 1 d.f.).

Dependency between Competing Risks

- A well-known problem with the multinomial logit model is the ‘independence of irrelevant alternatives’ (IIA) assumption
- In the context of competing risks, IIA implies that the probability of one event relative to ‘no event’ is independent of the probabilities of each of the other events relative to ‘no event’
- This may be unreasonable if some types of event can be regarded as similar
- Note that the multivariate binary model makes the same assumption

Dependency between Competing Risks: Example

Suppose we wish to study partnership formation: transitions from single (S) to marriage (M) or to cohabitation (C).

- Under IIA, assume probability of C vs. S is uncorrelated with probability of M vs. S
- E.g. if there is something unobserved (not in $\mathbf{x}$) that made M infeasible, we assume those who would have married distribute themselves between C and S in the same proportions as those who originally chose not to marry
- But as M and C are similar, 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*)

Relaxing the Independence Assumption

- Including individual-specific random effects allows for dependence due to time-invariant individual characteristics (e.g. attitudes towards marriage/cohabitation)

- But it does not allow for unmeasured factors that vary across episodes (e.g. marriage is not an option if respondent or their partner is already married)

Modelling Transitions between More than 2 States

So far we have considered (i) transitions between two states, and (ii) transitions from a single state with multiple destinations.

We can bring these together in a general model, allowing for different destinations from each state.

Example: partnership transitions

- Formation: S $\rightarrow$ M, S $\rightarrow$ C
- Conversion of C to M (same partner)
- Dissolution: M $\rightarrow$ S, C $\rightarrow$ S (or straight to new partnership)

Estimate 5 equations simultaneously (with correlated random effects)

Example of Multiple States with Competing Risks

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

- Estimate 3 equations jointly: binary logit for nonuse $\rightarrow$ use, and multinomial logit for transitions from use

- Details in Steele et al. (2004) *Statistical Modelling*

Selected Results: Coefficients and SEs

	Use → nonuse (Discontinuation)		Use → new method (Method switch)		Nonuse → use	
Urban	0.13	(0.04)	0.06	(0.05)	0.26	(0.04)
SES						
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 Alternative Models

	Discontinuation	Method switch	Nonuse $\to$ use
Discontinuation	1		
Method switch	0.020	1	
	0.011		
Nonuse $\to$ use	-0.783*	0.165*	1
	-0.052	0.095	

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)

Software for Competing Risks

- Multivariate binary response model can be framed as a multilevel random coefficients model
  - Coefficients for the response dummies are intercepts which vary randomly across individuals to give random effects for each type of event
  - Software options as for one-state recurrent events model, but using `xtmelogit` in Stata
- Multinomial model cannot currently be fitted in Stata (apart from via `runmlwin`). Other options include SAS (`PROC NLMIXED`), MLwiN and aML

5. Multiprocess models

Endogeneity in a 2-Level Continuous Response Model

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

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

where $\mathbf{x}_{ij}$ is a set of covariates with coefficients β, u_j is the level 2 random effect (residual) $\sim N(0, \sigma_u^2)$ and e_{ij} is the level 1 residual $\sim N(0, \sigma_e^2)$.

One assumption of the model is that $\mathbf{x}_{ij}$ is uncorrelated with both u_j and e_{ij}, i.e. we assume that $\mathbf{x}_{ij}$ is exogenous.

This may be too strong an assumption. If unmeasured variables affecting y_{ij} also affect one or more covariates, then those covariates will be endogenous.

2-Level Endogeneity: Example

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

Some of the factors that influence birth weight may also influence 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 attitudes/behaviour). Some of these may be unobserved.

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

This will lead to correlation between z and u and/or e and, if ignored, a biased estimate of the coefficient of z and possibly covariates correlated with z.

Illustration of Impact of Endogeneity at Level 1

Suppose the 'true' effect of z_{ij} on y_{ij} is positive, i.e. more antenatal visits is associated with a higher birth weight.

Suppose that w_{ij} is 'difficulty of pregnancy'. We would expect $\text{corr}(w, y) < 0$, and $\text{corr}(w, z) > 0$.

If w is unmeasured it is absorbed into e, leading to $\text{corr}(z, e) < 0$.

If we ignore $\text{corr}(z, e) < 0$, the estimated effect of z on y will be biased downwards.

The disproportionate presence of high w women among those getting more antenatal care (high z) suppresses the positive effect of z on y.

Illustration of Impact of Endogeneity at Level 2

As before, suppose the 'true' effect of z on y is positive, i.e. more antenatal visits is associated with a higher birth weight.

Suppose that w_j is 'healthcare knowledge' which is constant across the observation period. We would expect $\text{corr}(w, y) > 0$, and $\text{corr}(w, z) > 0$.

If w is unmeasured it is absorbed into u, leading to $\text{corr}(z, u) > 0$.

Question: What effect would ignoring $\text{corr}(z, u) > 0$ have on the estimated effect of z on y?

Handling Endogeneity in a Single-Level Model

To fix ideas, we will start with the simplest case: outcome y and endogenous predictor z both continuous.

E.g. y_i birth weight of last born child of woman i, z_i number of antenatal visits.

We specify a simultaneous equations model (SEM) for z and y:

$$\begin{aligned} z_i &= \beta^z \mathbf{x}_i^z + e_i^z \\ y_i &= \beta^y \mathbf{x}_i^y + \gamma z_i + e_i^y \end{aligned}$$

where $\mathbf{x}_i^z$ and $\mathbf{x}_i^y$ are exogenous covariates (assumed to be uncorrelated with e_i^z and e_i^y).

Estimation

If $\mathrm{corr}(e_i^z, e_i^y) = 0$, OLS of the equation for y_i is optimal.

Endogeneity of z_i will lead to $\mathrm{corr}(e_i^z, e_i^y) \neq 0$ and an alternative estimation procedure is required. The most widely used approaches are:

- 2-stage least squares (2SLS)
- Joint estimation of equations for z and y (Full Information Maximum Likelihood, FIML)

Estimation: 2-Stage Least Squares

1. OLS estimation of equation for z_i and compute $\hat{z}_i = \hat{\beta}^z \mathbf{x}_i^z$

2. OLS estimation of equation for y_i replacing z_i by prediction $\hat{z}_i$

3. Adjust standard errors in (2) to allow for uncertainty in estimation of $\hat{z}_i$

Idea: $\hat{z}_i$ is 'purged' of the correlated unobservables e_i^z, so $\hat{z}_i$ uncorrelated with e_i^y.

Estimation: FIML

Treat z_i and y_i as a bivariate response and estimate equations jointly.

Usually assume e_i^z and e_i^y follow a bivariate normal distribution with correlation ρ_e^{zy}.

- Can be estimated in a number of software packages (e.g. `mvreg` in Stata or Sabre)
- Sign of $\hat{\rho}_e^{zy}$ signals direction of bias
- Generalises to mixed response types (e.g. binary z and duration y)
- Generalises to clustered data (multilevel multivariate model)

Testing for Exogeneity of z

To test the null hypothesis that z is exogenous:

2SLS

Estimate $y_i = \beta^y \mathbf{x}_i^y + \gamma z_i + \delta \hat{e}_i^z + e_i^y$ via OLS

where $\hat{e}_i^z$ is the estimated residual from fitting the 1st stage equation for z_i

Test $\text{H}_0 : \delta = 0$ using t (or Z) test.

FIML

Jointly estimate equations for z_i and y_i to get estimate of residual correlation ρ_e^{zy}.

Test $\text{H}_0 : \rho_e^{zy} = 0$ using likelihood ratio test.

Identification

Whatever estimation approach is used, identification of the simultaneous equations model for z and y requires covariate exclusion restrictions.

$\mathbf{x}_i^z$ should contain at least one variable that is not in $\mathbf{x}_i^y$.

In our birth weight example, need to find variable(s) that predict antenatal visits (z) but not birth weight (y).

Call such variables instruments.

Note: The term 'IV estimation' is commonly used interchangeably with 2SLS, but both methods require instruments.

Requirements of an Instrument (1)

Need to be able to justify, on theoretical grounds, that the instrument affects z but not y (after controlling for z and other covariates).

E.g. indicator of access to antenatal care may be suitable instrument for no. visits, but only if services are allocated randomly (rare). Instruments can be very difficult to find.

If there is > 1 instrument, the model is said to be over-identified.

Requirements of an Instrument (1)

Testing over-identifying restrictions

Instruments should not affect y after controlling for z.

Fit the SEM with all but one instrument in the equation for y and carry out a joint significance test of the included instruments. If the restrictions are valid, they should not have significant effects on y.

Instruments should be correlated with z

Carry out joint significance test of effects of instruments on z.

Also check how well instruments (together with other covariates) predict z. Bollen et al. (1995) suggest a simple probit for y is preferred if $R^2 < 0.1$.

Effect of Fertility Desires on Contraceptive Use (1)

Reference: Bollen, Guilkey and Mroz (1995), *Demography*.

Interested in the impact of number of additional children desired (z, continuous) on use of contraception (y, binary).

Unmeasured variable affecting both z and y could be 'perceived fecundity'.

Women who believe they have low chance of having a(nother) child may lower fertility desires and not use contraception $\rightarrow$ $\text{corr}(e_i^z, e_i^y) > 0$.

Effect of Fertility Desires on Contraceptive Use (2)

Expect 'true' effect of fertility desires (z) on contraceptive use (y) to be negative.

- If residual correlation ignored, negative effect of z on y will be understated (may even estimate a positive effect)

- Estimated effects of covariates correlated with z also biased (e.g. whether heard family planning message)

Effect of Fertility Desires on Contraceptive Use (3)

Cross-sectional data: z and y refer to time of survey.

Use 2SLS: OLS for z equation, probit for y.

Instruments: Indicators of health care facilities in community when woman was age 20 (supplementary data).

Results:

- Residual correlation estimated as 0.07
- Stronger negative effect of z after allowing for endogeneity (changes from -0.17 to -0.28), but large increase in SE
- But fail to reject null that z is exogenous, so simple probit for contraceptive use is preferred

Handling Endogeneity in a Multilevel Model

Let's return to the multilevel case with y_{ij} the birth weight of child i of woman j, z_{ij} number of antenatal visits.

We specify a multilevel simultaneous equations (multiprocess) model for z and y:

$$\begin{aligned} z_{ij} &= \beta^z \mathbf{x}^z_{ij} + u^z_j + e^z_{ij} \\ y_{ij} &= \beta^y \mathbf{x}^y_{ij} + \gamma z_{ij} + u^y_j + e^y_{ij} \end{aligned}$$

where u^z_j and u^y_j are normally distributed woman-level random effects, and $\mathbf{x}^z_{ij}$ and $\mathbf{x}^y_{ij}$ are exogenous covariates (assumed to be uncorrelated with u^z_j, u^y_j, e^z_{ij} and e^y_{ij}).

Estimation

If $\text{corr}(u_j^z, u_j^y) = 0$ and $\text{corr}(e_{ij}^z, e_{ij}^y) = 0$, the equation for y_{ij} can be estimated as a standard multilevel model.

However, endogeneity of z_{ij} will lead to $\text{corr}(u_j^z, u_j^y) \neq 0$ or $\text{corr}(e_{ij}^z, e_{ij}^y) \neq 0$ (or both).

If z_{ij} is endogenous we need to estimate equations for z and y jointly.

In the most general model, we assume $(u_j^z, u_j^y) \sim$ bivariate normal and $(e_{ij}^z, e_{ij}^y) \sim$ bivariate normal. The SEM is a multilevel bivariate response model.

Identification (1)

Identification of the full multilevel SEM for z and y, with $\text{corr}(u_j^z, u_j^y) \neq 0$ and $\text{corr}(e_{ij}^z, e_{ij}^y) \neq 0$, requires covariate exclusion restrictions:

$\mathbf{x}_{ij}^z$ should contain at least one variable (an instrument) that is not in $\mathbf{x}_{ij}^y$.

e.g. need to find variable(s) that predict antenatal visits (z) but not birth weight (y).

Call such variables instruments.

BUT if one of the residual covariances is assumed equal to zero, covariate exclusions are not strictly necessary for identification.

Identification (2)

Suppose we are prepared to assume that endogeneity of z is due to a residual correlation at the woman level but not at the pregnancy level, i.e.

$\text{corr}(u_j^z, u_j^y) \neq 0$ but $\text{corr}(e_{ij}^z, e_{ij}^y) = 0$.

We are then assuming that bias in the estimated effect of number of antenatal visits on birth weight is due to selection on unmeasured maternal characteristics that are fixed across pregnancies.

Identification (3)

Given the difficulty in finding instruments, allowing only for selection on time-invariant unobservables (in a longitudinal design) is a common identification strategy BUT:

- It does not allow for selection on time-varying unobservables so some bias may be remain
- Some within-individual variation in z and y is required because we are estimating the effect of a change in z on y for a given woman (i.e. conditioning on woman-specific unobservables)

Allowing for Endogeneity in an Event History Model

Suppose that y_{ij} is the duration of episode i of individual j and z_{ij} is an endogenous variable. We first consider case where z is continuous and measured at the episode level.

We can extend our earlier recurrent events model to a SEM:

$$\begin{aligned} z_{ij} &= \beta^z \mathbf{x}^z_{ij} + u^z_j + e^z_{ij} \\ \log\left(\frac{p_{tij}}{1-p_{tij}}\right) &= \alpha^y \mathbf{D}^y_{tij} + \beta^y \mathbf{x}^y_{tij} + \gamma z_{ij} + u^y_j \end{aligned}$$

where p_{tij} is the probability of an event during interval t, $\alpha^y \mathbf{D}^y_{tij}$ is the baseline hazard, and $\mathbf{x}^y_{tij}$ a vector of exogenous covariates.

We assume $(u^z_j, u^y_j) \sim$ bivariate normal, i.e. we allow for selection on time-invariant individual characteristics.

Examples of Multilevel SEM for Event History Data

More generally z can be categorical and can be defined at any level (e.g. time-varying or a time-invariant individual characteristic).

We will consider two published examples before returning to our analysis of women's employment transitions:

- The effect of premarital cohabitation (z) on subsequent marital dissolution (y)
  - Lillard, Brien & Waite (1995), *Demography*
- The effect of access to family planning (z) on fertility (y)
  - Angeles, Guilkey & Mroz (1998), *Journal of the American Statistical Association*

Example 1: Premarital Cohabitation and Divorce

Couples who live together before marriage appear to have an increased risk of divorce.

Is this a 'causal' effect of premarital cohabitation or due to self-selection of more divorce-prone individuals into premarital cohabitation?

The analysis uses longitudinal data so observe women in multiple marriages (episodes). For each marriage define 2 equations:

- A probit model for premarital cohabitation (z)
- A (continuous-time) event history model for marital dissolution (y)

Each equation has a woman-specific random effect, u_j^z and u_j^y, which are allowed to be correlated

Premarital Cohabitation and Divorce: Identification

Lillard et al. argue that exclusion restrictions are unnecessary because of ‘within-person replication’.

Nevertheless they include some variables in the cohabitation equation that are not in the dissolution equation:

- Education level of woman’s parents
- Rental prices and median home value in state
- Sex ratio (indicator of ‘marriageable men/women’)

They examine the robustness of their conclusions to omitting these variables from the model.

Premarital Cohabitation and Divorce: Results (1)

Correlation between woman-specific random effects for cohabitation and dissolution estimated as 0.36.

Test statistic from a likelihood ratio test of the null hypothesis that $\text{corr}(u_j^z, u_j^y) = 0$ is 4.6 on 1 d.f. which is significant at the 5% level.

"There are unobserved differences across individuals which make those who are most likely to cohabit before any marriage also most likely to end any marriage they enter."

Premarital Cohabitation and Divorce: Results (2)

What is the impact of ignoring this residual correlation, and assuming premarital cohabitation is exogenous?

Estimated effect of cohabitation on log-hazard of dissolution

0.37 and strongly significant if $\text{corr}(u_j^z, u_j^y) = 0$ assumed

-0.01 and non-significant if $\text{corr}(u_j^z, u_j^y)$ allowed to be non-zero

Conclude that, after allowing for selection, there is no association between premarital cohabitation and marital dissolution.

Example 2: Access to Family Planning and Fertility

Does availability of family planning (FP) services lead to a reduction in fertility in Tanzania?

Problem: FP clinics are unlikely to be placed at random. They are likely to be targeted towards areas of greatest need, the type of area with high fertility.

Question: If true impact of access to FP is to increase birth spacing, how will ignoring targeted placement affect estimates of the impact?

Data and Measures

Birth histories collected retrospectively in 1992. Women nested within communities, so have a 3-level structure: births (level 1), women (level 2), communities (level 3).

Constructed woman-year file for period 1970-1991 with y_{tij}=1 if woman i in community j gave birth in year t. (Could have extra subscript for birth interval as we model duration since last birth.)

Community survey on services conducted in 1994. Construct indicators of distance to hospital, health centre etc. in year t, z_{tj}. Time-varying indicators derived from information on timing of facility placement.

Multiprocess Model for Programme Placement and Fertility

The model consists of 4 equations:

- Discrete-time event history model for probability of a birth in year t with woman and community random effects
- Logit models for placement of 3 types of FP facility in community j in year t with community random effects

Allow for correlation between community random effects for fertility and FP clinic placement.

Rather than assume normality, the random effects distribution is approximated by a step function using a 'discrete factor' method.

Programme Placement and Fertility: Identification

The following time-varying variables are included in the FP placement equations but not the fertility equation:

- National government expenditure on health
- Regional government expenditure on health
- District population as fraction of national population

These are based on time series data at the national, regional and community levels.

Programme Placement and Fertility: Findings

- From simple analysis (ignoring endogeneity of programme placement) find hospitals have more impact on reducing fertility than health centres
- But this analysis overstates impact of hospitals and understates effects of health centres
- Controlling for endogenous programme placement reveals that health centres have more impact than hospitals
- After controlling for endogenity, impact of FP facilities was 45% larger than in simpler analysis

Modelling Correlated Event Processes

Now suppose that z_{tij} is a time-varying endogenous predictor.

z_{tij} is often the outcome of a related event process.

Example: Marital dissolution and fertility

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

z_{tij} is number of children from marriage i of woman j at time t, the outcome of the birth history by t

See Lillard (1993) and Lillard & Waite (1993).

Multiprocess Model SEM for 2 Interdependent Events

Simultaneous discrete-time event history equations:

$$\begin{aligned}\text{logit}(p^z_{tij}) &= \alpha^z \mathbf{D}^z_{tij} + \beta^z \mathbf{x}^z_{tij} + u^z_j \\ \text{logit}(p^y_{tij}) &= \alpha^y \mathbf{D}^y_{tij} + \beta^y \mathbf{x}^y_{tij} + \gamma z_{ij} + u^y_j\end{aligned}$$

We assume $(u^z_j, u^y_j) \sim$ bivariate normal, i.e. we allow for selection on time-invariant individual characteristics.

The model can be extended to include outcomes of the y process in the model for z.

Example: Marital Dissolution and Fertility

Lillard's model has 2 (continuous-time) event history equations for:

- hazard of conception (leading to a live birth) at time t of marriage i of woman j
- hazard that marriage i of woman j ends at time t

Consider dummies for z_{tij}, the number of children from marriage i, in dissolution equation.

Marital Dissolution and Fertility: Results (1)

Lillard (1993) finds that the residual correlation between hazard of dissolution and hazard of a conception is estimated as -0.75 (se=0.20).

$\Rightarrow$ women with a below-average risk of dissolution ($u_j^y < 0$) tend to have an above-average chance of a marital conception ($u_j^z > 0$).

$\Rightarrow$ selection of women with a low dissolution risk into having children.

Question: If the 'true' effect of having children is to reduce the risk of dissolution, what impact would this type of selection have on estimates of this effect?

Marital Dissolution and Fertility: Results (2)

Estimated effects (SE) of number of children from current marriage on log-hazard of dissolution before and after accounting for residual correlation:

# children	Before		After	
0 (ref)	0		0	
1	−0.56	(0.10)	−0.33	(0.11)
2+	−0.01	(0.05)	0.27	(0.07)

Selection of low dissolution risk women into categories 1 and 2+

Other Examples of Correlated Event Histories

- Employment transitions and fertility (next example)
- Partnership formation and employment
- Residential mobility and fertility
- Residential mobility and employment
- Residential mobility and partnership formation/dissolution

Multiprocess Model for Entry into Employment and Fertility (1)

At the start of the course (and in computer exercises) we fitted multilevel models for the transition from non-employment (NE) to employment (E) among British women.

Among the covariates was a set of time-varying fertility indicators:

- Due to give birth within next year
- Number of children aged ≤ 5 years
- Number of children aged > 5 years

These are outcomes of the fertility process which might be jointly determined with employment transitions.

Multiprocess Model for Entry into Employment and Fertility (2)

Denote by y_{tij}^{NE} and y_{tij}^{B} binary indicators for leaving non-employment and giving birth during year t.

Estimate 2 simultaneous equations (both with woman-specific random effects):

- Discrete-time logit for probability of a birth
- Discrete-time logit for probability of NE $\rightarrow$ E (with fertilty outcomes as predictors)

Note: While we could model births that occur during non-employment, it would be more natural to model the whole birth process (in both NE and E). In the following analysis, we consider *all* births.

Estimation of Multiprocess Model

We can view the discrete-time multiprocess model as a multilevel bivariate response model for the binary responses y_{tij}^{NE} and y_{tij}^{B}.

- Stack the employment and birth responses into a single response column and define an index r which indicates the response type (e.g. $r = 1$ for NE and $r = 2$ for B)

- Define dummies for r which we call r_1 and r_2 say

- Multiply r_1 and r_2 by the covariates to be included in the NE and B equations respectively

- Fit woman-level random effects to r_1 and r_2 and allow to be correlated

Entry into Employment and Fertility: Residual Correlation

Likelihood ratio test statistic for test of null hypothesis that $\text{corr}(u_j^{NE}, u_j^{B}) = 0$ is 8 on 1 d.f.

$\Rightarrow$ reject the null and choose the multiprocess model.

Correlation between woman-level random effects, u_j^{NE} and u_j^{B}, estimated as 0.34 (se=0.11).

The positive correlation implies that women whose unobserved characteristics are associated with a high probability of a birth (e.g. latent preference for childbearing) tend also to enter employment quickly after a spell of non-employment.

Effects of Fertility Outcomes on Entry into Employment

	Single process		Multiprocess	
Variable	Est.	(se)	Est.	(se)
Imminent birth (within 1 year)	−0.84*	(0.13)	−1.01*	(0.14)
No. children $\leq$ 5 yrs (ref=0)				
1 child	−0.21*	(0.10)	−0.35*	(0.11)
≥ 2	−0.35*	(0.14)	−0.60*	(0.17)
No. children $>$ 5 yrs (ref=0)				
1 child	0.25*	(0.12)	0.18	(0.12)
≥ 2	0.45*	(0.12)	0.27*	(0.13)

* $p < 0.5$

Selection of women with high NE $\rightarrow$ E probability into categories 1 and ≥ 2

Multiple States and Correlated Processes

We can extend the multiprocess model to include transitions between multiple states and further correlated processes.

E.g. we could model two-way transitions between NE and E jointly with births, leading to 3 simultaneous equations and 3 correlated random effects.

Stack employment transition and birth responses into a single column with a 3-category response indicator r (e.g. $r=1$ for employment episodes, $r=2$ for non-employment episodes, $r=3$ for birth intervals).

Create dummies for r and interact with covariates as for two-state and multiprocess models.

Employment Transitions and Fertility: Random Effects Correlation Matrix

	NE $\rightarrow$ E	E $\rightarrow$ NE	Birth
NE $\rightarrow$ E	1		
E $\rightarrow$ NE	0.62 (0.12)	1	
Birth	0.45 (0.11)	0.23 (0.08)	1

Standard errors in brackets

Effects of Fertility Outcomes on Exit from Employment

	Single process		Multiprocess	
Variable	Est.	(se)	Est.	(se)
Imminent birth (within 1 year)	2.31*	(0.14)	2.23*	(0.15)
No. children $\leq$ 5 yrs (ref=0)				
1 child	0.41*	(0.10)	0.31*	(0.11)
$\geq$ 2	0.33	(0.17)	0.15	(0.18)
No. children $>$ 5 yrs (ref=0)				
1 child	−0.35*	(0.12)	−0.34	(0.12)
$\geq$ 2	−0.28*	(0.12)	−0.37*	(0.13)

* $p < 0.5$

Selection of women with high NE $\rightarrow$ E probability into categories 1 and $\geq$ 2

Example: Family Disruption and Children's Education

Research questions:

- What is the association between disruption (due to divorce or paternal death) and children's education?
- Are the effects of disruption the same across different educational transitions?
- To what extent can the effect of divorce be explained by selection?
  - There may be unobserved factors affecting both parents' dissolution risk and their children's educational outcomes

Reference: Steele, Sigle-Rushton and Kravdal (2009), *Demography*.

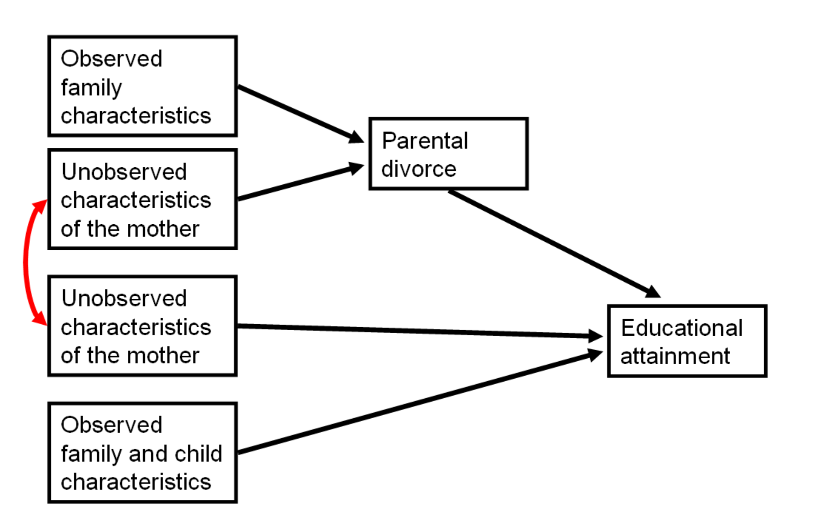
Strategies for Handling Selection

- Exploit longitudinal study designs
  - e.g. measures of child wellbeing before divorce, measures of parental conflict and family environment
- Exploit differences across space
  - compare children living in places with differences in availability of divorce (e.g. US states)
- Compare siblings
  - Siblings share parents (or parent) but may have different exposure to disruption
- Multiprocess (simultaneous equations) models

SEM for Parental Divorce and Children's Education

- Selection equation: event history model for duration of mother's marriage(s)
- Sequential probit model for children's educational transitions (nested within mother)
- Equations linked by allowing correlation between mother-specific random effects (unmeasured maternal characteristics)
- Estimated using aML software

Simple Conceptual Model



Sequential Probit Model for Educational Transitions (1)

- View educational qualifications as the result of 4 sequential transitions:
  - Compulsory to lower secondary
  - Lower to higher secondary (given reached lower sec.)
  - Higher secondary to Bachelor's (given higher sec.)
  - Bachelor's to postgraduate (given Bachelor's)
- Rather like a discrete-time event history model
- Advantages:
  - Allow effects of disruption to vary across transitions
  - Can include children who are too young to have made *all* transitions

Sequential Probit Model for Educational Transitions (2)

Transition from education level r for child i of woman j indicated $y_{ij}^{(r)}$ =1 if child attains level $r+1$ and 0 if stops at r.

$$y_{ij}^{(r)^*} = \beta^{(r)}\mathbf{x}_{ij} + \gamma^{(r)}\mathbf{z}_{ij} + \lambda^{(r)}u_j + e_{ij}^{(r)}, \qquad r = 1, \ldots, 4$$

$y_{ij}^{(r)^*}$ latent propensity underlying $y_{ij}^{(r)}$

$\mathbf{z}_{ij}$ potentially endogenous indicators of family disruption

$\mathbf{x}_{ij}$ child and mother background characteristics

u_j mother-specific random effect

$e_{ij}^{(r)}$ child and transition-specific residual

Unobserved Heterogeneity: Educational Transitions

Parameter	Transition	Estimate	(SE)
Random effect loading			
$\lambda^{(1)}$	To low secondary	$1^{\dagger}$	-
$\lambda^{(2)}$	To high secondary	1.078*	(0.041)
$\lambda^{(3)}$	To Bachelor's	0.864*	(0.041)
$\lambda^{(4)}$	To Master's +	0.584*	(0.077)
Standard deviation σ_u		0.498*	(0.014)

*Significant at 1% level; †Constrained to equal 1

Effect of mother-level unobservables less important for later transitions.

Evidence for Selection

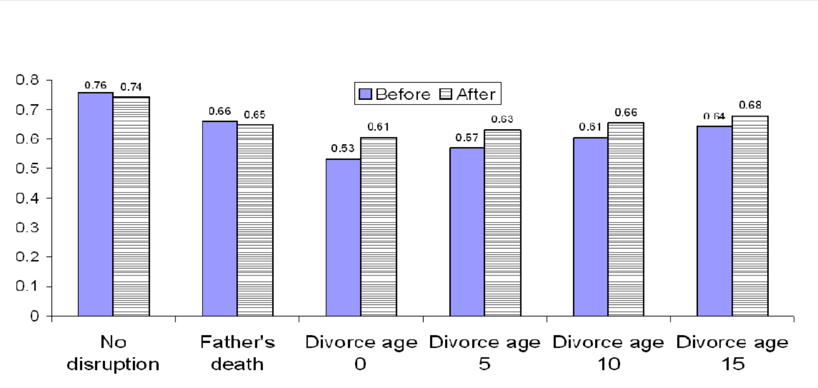
- Residual correlation between dissolution risk and probability of continuing in education estimated as -0.43 (se=0.02)
- Suggests mothers with above-average risk of divorce tend to have children with below-average chance of remaining education
- Note that we are controlling only for selection on unobservables at the mother level (i.e. fixed across time)

Effects of Disruption on Transitions in Secondary School

	Compulsory to lower secondary		Lower to higher secondary	
	Model 1 ($\rho_{uv} = 0$)	Model 2 ($\rho_{uv} \neq 0$)	Model 1 ($\rho_{uv} = 0$)	Model 2 ($\rho_{uv} \neq 0$)
Parents separated	-0.580*	-0.349*	-0.631*	-0.386*
Age at separation	0.019*	0.013*	0.018*	0.013*
Father died	-0.201*	-0.178*	-0.318*	-0.295*
Female	0.217*	0.217*	0.318*	0.320*
Female×separation	0.034	0.034	0.005	0.003
Female×father died	-0.126	-0.125	0.107	0.110

*Significant at 1% level

Predicted Probabilities of Continuing Beyond Lower Secondary (Before and After Allowing for Selection)



Software for multiprocess modelling

- Stata
  - Responses of same type only: continuous (`xtmixed`) or binary (`xtmelogit`)
  - Can handle multiple processes in theory, but slow
- MLwiN (and `runmlwin` in Stata)
  - Designed for multilevel modelling; multiple levels
  - Handles mixtures of continuous and binary responses
  - Markov chain Monte Carlo estimation
- Sabre
  - Developed for analysis of recurrent events
  - Handles mixtures of response types; up to 3 processes; 2 levels
- aML
  - Designed specifically for multilevel multiprocess modelling
  - Mixtures of response types; multiple processes and levels

Further Reading

Bibliography: Discrete-time Methods & Recurrent Events

Singer, J. D., and Willett, J. B. (2003) *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Davis, R.B., Elias, P. and Penn, R. (1992) The relationship between a husbands unemployment and a wifes 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: Multiple States & 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.

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.

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

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. *Statistical Modelling*, 4: 145-159.

Bibliography: Multiprocess Modelling (1)

Angeles, G., Guilkey, D.K. and Mroz, T.A. (1998) Purposive program placement and the estimation of family planning program effects in Tanzania. *Journal of the American Statistical Association*, 93: 884-899.

Bollen, K.A., Guilkey, D.K. and Mroz, T.A. (1995) Binary outcomes and endogenous explanatory variables: tests and solutions with an application to the demand for contraceptive use in Tunisia. *Demography*, 32: 111-131.

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

Lillard, L.A., Brien, M.J. and Waite, L.J. (1995) Premarital cohabitation and subsequent marital dissolution: a matter of self-selection? *Demography*, 32: 437-457.

Bibliography: Multiprocess Modelling (2)

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.

Steele, F., Sigle-Rushton, W. and Kravdal, Ø. (2009) Consequences of family disruption on children's educational outcomes in Norway. *Demography*, 46: 553-574.

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.