EXERCISE 3: MULTILEVEL MODELS FOR MULTIPLE STATES, USING MLwiN

In this exercise, we will be using *MLwiN* to fit multilevel discrete-time event history models for transitions from multiple states. At the same time we will allow for unobserved heterogeneity by including individual-level random effects.

We will model partnership dissolution, treating marriage and cohabitation as two different states. For the purposes of this exercise, an episode is defined as a continuous period spent with the same partner. A respondent may move from cohabitation to marriage within the same episode, but our end event of interest is dissolution, defined as the end of co-residential living. (In the next exercise, we will model the transition from cohabitation to marriage.)

1. Data

The data consist of 1000 women from the National Child Development Study (NCDS). For this subsample, we consider all partnerships between the ages of 16 and 42. We include individual-level random effects to allow for correlation between the durations of partnerships from the same woman. These women contribute 1345 partnerships (counting marriage immediately preceded by cohabitation with the same partner as a single partnership); within these partnerships are 675 cohabitation and 1065 marriage episodes.

The data are already in discrete-time format, with one record per six-month time interval spent in a partnership. We will adjust for different lengths of exposure to the risk of separation within a six-month interval. The 1000 women contribute a total of 33,212 six-monthly observations. The ascii dataset **ex3.dat** contains the following variables:

WOMAN	Woman identifier
STATE	State indicator (1=cohabitation, 2=marriage)
SEP	Event indicator (1=separation, 0=still together)
ATRISK	Number of months of 6-month interval for which woman exposed to risk of separation
DUR	Partnership duration, time-varying (in 6-month intervals)
PRMAR	Previously married (1=yes, 0=no)
PRCOH_ANY	Ever previously cohabited (1=yes, 0=no)
PRCOH	Previous cohabitation (1=never, 2=with current partner only, 3=with previous partner only, 4=with both current and previous partners)
AGESTART	Age at start of partnership (1=less than 20, 2=20-24, 3=25-29, 4=30+)
EDUC	Number of years of post-16 education, time-varying (1=none, 2=1-2 years, 3=3-5 years, 4=6+ years)
NKIDY	Presence of preschool children, time-varying (0=none, 1=1+)

NKIDO	Presence of older children, time-varying (0=none, 1=1+)

2. Data Preparation and Setting up a Simple Model

The file **ex3_macro.txt** is a *MLwiN* macro. This macro contains syntax to read the ascii file into *MLwiN*, create new variables (including dummies for the marriage and cohabitation states, and their interactions with partnership duration and covariates), and set up a simple multistate model.

Note: In the current version of *MLwiN*, while it is possible to specify binary response models through the Equations window (see Chapter 9 of the User Guide), there is no single command for specifying the model using syntax. Therefore, for convenience, we specify a multinomial logit model, of which the binary logit model is a special case.

The contents of the macro file are given at the end of this handout (with line numbers added for easy reference). Lines 2-3 read in the ascii data file to columns C1-C12 of the *MLwiN* worksheet. These columns are named (lines 4-5). A number of new variables are created: (1) squared duration (the hazard of dissolution will be modelled as a quadratic function of partnership duration), (2) a variable which will be the level 1 identifier (actually any variable could be used for this - it is the woman identifier which is important here). In addition, dummy variables indicating the current partnership state are created. These are both multiplied by the duration variables and age at start of the partnership to allow for state-specific effects. The two age variables (the interactions with age and the state dummies) are declared as categorical and labels are attached to the four categories (lines 29-30). When age is added to the model as a covariate, dummy variables will be created for each category, taking the first category as the reference by default.

The syntax in lines 35 onwards specify a simple single-level multistate model. Because we are fitting the model as a multinomial logit model, we first declare the binary response as categorical and take the category coded zero as the reference. The next step is to set up the multilevel structure (even though we start with a single-level model). We then convert the binary response to a proportion which takes into account the number of months of exposure within each six-month interval. The **atrisk** variable is declared as the denominator for this proportion. Finally we add the state dummies and their interactions with duration, duration-squared and age as explanatory variables.

To open and run the macro in *MLwiN*:

From the **File** menu, select **Open Macro**

Locate the file **ex3_macro.txt**

Click on the **Execute** tab to run the macro

3. Fitting a Simple Multistate Model

To run the single-level model which ignores unobserved heterogeneity:

Open the **Equations** window (**Model → Equations**)

Click **Nonlinear** and select **Use Defaults**

Click **Start**

Click on **Estimates** to see the estimates

You should get the following results:

Notice that the suffix **.sep** has been added to all variables. This is because we have specified a multinomial model. The suffix reminds us that we are contrasting the hazard of separation with the hazard of continuing the partnership. The coefficient of **c.sep** is the intercept in the equation for dissolution from cohabitation, while the coefficient of **m.sep** is the intercept for marital dissolution. Similarly, the coefficients of variables prefixed by **c** are effects on the log-odds of separation from cohabitation.

What is the effect of age on separation from the two partnership states?

4. Hypothesis Testing

To test for age effects on the risk of separation from a given state, we can carry out a Normal (or t) test in the usual way. For example, to test for a difference in the risk of marital separation between women aged 20-24 when they first lived with their partner and those who were teenagers (the reference category) we can simply divide the estimate of **m_age20-24.sep** by its standard error and compare to a Normal distribution. What do you conclude?

We might also be interested in testing whether this difference between age groups 20-24 and < 20 years is the same for married and cohabiting women. The null hypothesis for this test is that the coefficients of **c_age20-24.sep** and **m_age20-24.sep** are equal, i.e. that $\beta_3 - \beta_9 = 0$.

While we could test this hypothesis using a Normal test, it is more convenient in *MLwiN* to use a Wald test.

From the **Model** menu, select **Intervals and tests**

At the bottom of the window check **fixed**

Type **1** next to **fixed: c_age20-24.sep** and **–1** next to **fixed: m_age20-24.sep**

Click **Calc**

The **Intervals and tests** window should look like this:

Intervals and tests	# 1
fixed : c.sep	0.000
fixed : c_dur.sep	0.000
fixed : c_dursq.sep	0.000
fixed : c_age20-24.sep	1.000
fixed : c_age25-29.sep	0.000
fixed : c_age30+.sep	0.000
fixed : m.sep	0.000
fixed : m_dur.sep	0.000
fixed : m_dursq.sep	0.000
fixed : m_age20-24.sep	-1.000
fixed : m_age25-29.sep	0.000
fixed : m_age30+.sep	0.000
constant(k)	0.000
function result(f)	0.561
f-k	0.561
chi sq, (f-k)=0. (1df)	3.612
+/- 95% sep.	0.578
+/- 95% joint	0.578

joint chi sq test(1df) = 3.612

○ random ◉ fixed # of functions 1   Calc   Help

The test statistic is 3.6 which when compared to a chi-squared distribution on 1 d.f. gives a p-value which is slightly above 0.05 (you can do this using **Basic Statistics → Tail Areas**), so this is some evidence that the effect on separation of partnering in one's early 20s (relative to one's teenage years) depends on the type of partnership.

We can also test for an overall effect of age on either transition using a joint Wald test. For example, suppose we wish to test whether there is an age effect on the risk of marital separation. The null hypothesis is that $\beta_9 = \beta_{10} = \beta_{11} = 0$.

From the **Model** menu, select **Intervals and tests**

At the bottom of the window check **fixed**

> Change **# of functions** to **3**
>
> Type a **1** next to each of **fixed: m_age20-24.sep**, **m_age25-29.sep** and **m_age30+.sep**
>
> Click **Calc**

The test statistic is labelled **joint chi sq test**. What do you conclude?

5. Allowing for Unobserved Heterogeneity

To add random effects for each state, we need to allow the coefficients of the variables **c.sep** and **m.sep** to vary randomly across women.

> In the **Equations** window click on **c.sep** (or its coefficient) and check **k(woman_long)**, then **Done**
>
> Repeat for **m.sep**

The **Equations** window should look like this (you may need to click **Estimate**s and/or **+**):

The between-woman variance in the log-odds of separation is σ_{v0}^2 for cohabitation, and σ_{v6}^2 for marriage. The covariance between the state-specific random effects is σ_{v06}. If women with a high propensity to separate from cohabitation also have a high propensity to separate from marriage, we would expect a positive covariance estimate.

We will start by estimating the model using quasi-likelihood procedures, but as these are unreliable for small clusters we will switch to MCMC methods after convergence. It turns out that we run into estimation problems if we introduce all three parameters in the random effect covariance matrix at the same time. We will therefore first remove the covariance term, and then add it in again upon convergence of the simplified model. It is always a good

idea to build complex multilevel models gradually.

Click on σ_{v06} and respond 'Yes' to the question about removing this term. The term is replaced by 0, indicating that this parameter has been constrained to equal zero

Click on **More** to fit the model

On convergence, click on the 0 which replaced σ_{v06} and again respond 'Yes' to the question about adding this term

Click on **More** to fit this model

On convergence, click **Estimation Control** then **MCMC**

We will run MCMC for just a short time for illustration, so edit **Monitoring Chain Length** to **1000**

Click **Done** to activate the MCMC settings, then **Start**

MCMC will take a few minutes to run. In the meantime you may wish to start looking at the exercises below.

You should get the following results:
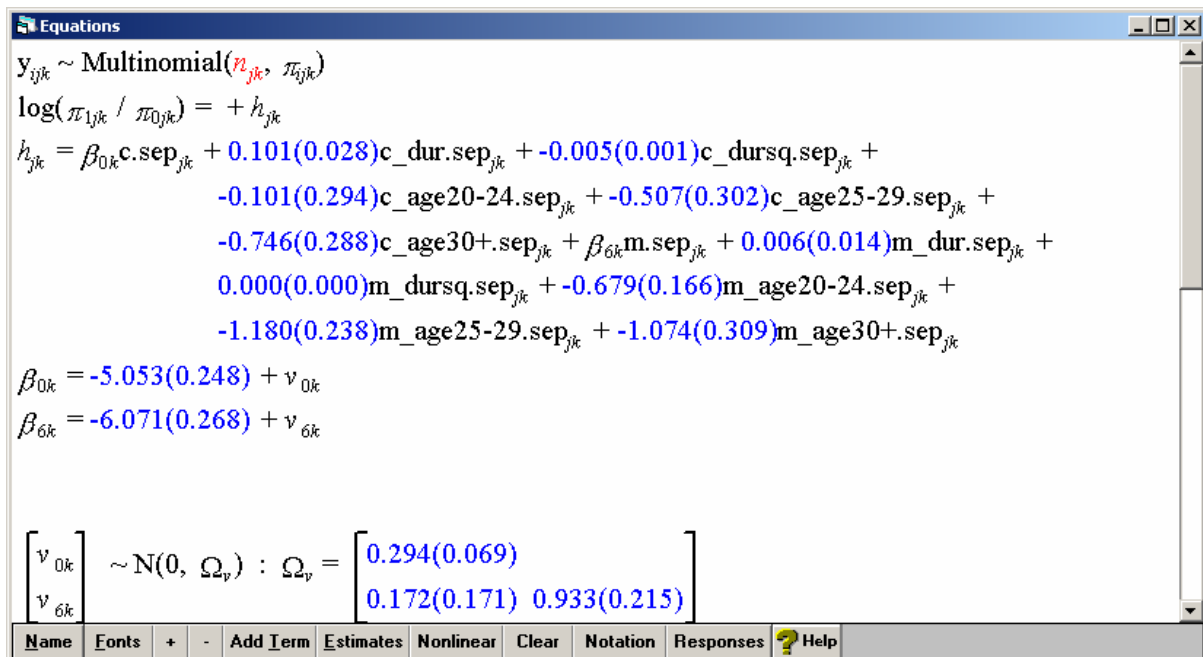
To see the *correlation* between the random effects:

From the **Model** menu, select **Estimate tables**
Change from **FIXED PART** to **level 3: woman_long**
Check the box under **C**

The correlation is estimated as 0.329.

Before interpreting the results, we should check whether convergence has been achieved. A number of criteria have been proposed for assessing convergence in MCMC estimation (see Browne, W.J. (2003) *MCMC Estimation in MLwiN v2.0* for further details). One approach is to inspect the trajectories of each parameter. We will start by looking at the trajectory of the random effect covariance parameter:

From the **Model** menu, select **Trajectories** and respond **OK** to the prompt that appears

Click on **Select** at the bottom of the window and then **woman_long: m.sep/c.sep** (the covariance between the random effects for the marriage and cohabitation intercepts at the woman level) followed by **Done**

We will look at the estimates at each of the 1000 iterations, so next to **view last** select **1000**

You should obtain the following plot:
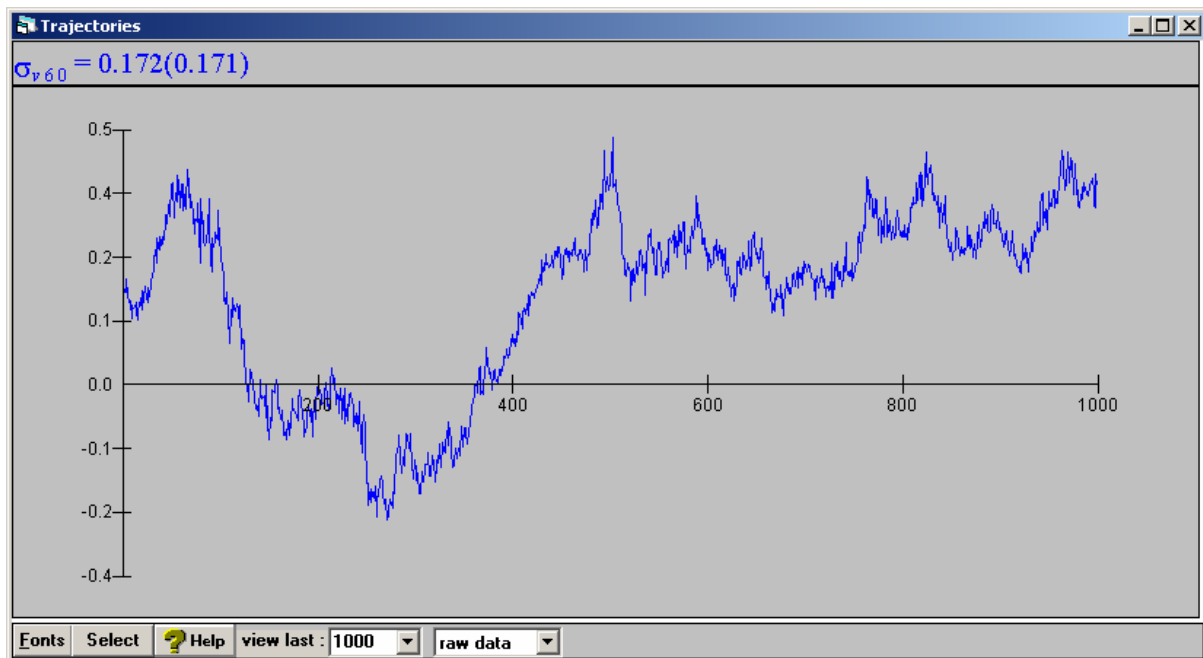
Notice that near the start of the chain there is a sharp decrease in the estimates followed by an equally sharp increase, before the chain begins to settle down. This is an example of poor mixing and suggests that we need a longer burn-in (the part of the chain we throw away). We will also increase the chain length. A chain with good mixing should not show any trend like this; the estimates should fluctuate randomly. If we increase the burn-in to 5000 and the monitoring chain length to 50,000, we obtain the following results (after several hours):

Note that, as anticipated, the estimated covariance between the random effects is positive. (The correlation is estimated as 0.498.) We can obtain interval estimates for each parameter, based on the 50,000 samples in the chain. For example, a 95% interval estimate is given by the 2.5% and 97.5% points of the simulated parameter distribution. To see the interval estimates and other summary statistics, as well as further convergence diagnostics, we click on a trajectories plot. If we do this for the covariance parameter, we get a 95% interval estimate of (-0.05, 0.75). As this includes zero, the covariance is not significant at the 5% level.

6. Exercises

Because of the length of time required to estimate models using MCMC, we will ignore unobserved heterogeneity for now. In practice, it is advisable to carry out preliminary analysis using single-level models, fitting only selected models with random effects. Alternatively, the estimates obtained using quasi-likelihood methods are useful as an approximation.

To remove the random effects:

Click **Estimation Control** then **IGLS/RIGLS**, then **Done**
Click on **c.sep** and deselect **k(woman_long)**
Repeat for **m.sep**. The woman-level random effect covariance matrix should disappear
Click **Start** to refit the single-level model

- Modify **ex3_macro.txt** to add in other covariates to the model. Remember to declare variables with more than two categories as categorical. All covariates will need to be multiplied by **c** and **m** to allow their effects on separation to vary for cohabitation and marriage.

- Do the effects of other covariates differ for cohabitation and marriage?

```
1    NOTE: Exercise 3 - macro to set up multiple state model in MLwiN

2    dinp c1-c12
3    ex3.dat

4    name c1 'woman' c2 'state' c3 'sep' c4 'atrisk' c5 'dur' c6 'prmar' c7 'prcoh_any'
5    name c8 'prcoh' c9 'agestart' c10 'educ' c11 'nkidy' c12 'nkido'

6    NOTE: Calculate duration-squared variable
7    calc c13='dur'*'dur'
     name c13 'dursq'

8    NOTE: Create level 1 ID (coded 1,2, . . ., 33212)
9    code 33212 1 1 c14
10   name c14 'lev1id'

11   NOTE: Create state indicators
12   calc c16=('state'==1)
13   calc c17=1-c16
14   name c16 'c' c17 'm'

15   NOTE: Calculate interactions between state indicators and duration
16   calc c18='c'*'dur'
17   calc c19='m'*'dur'
18   calc c20='c'*'dursq'
19   calc c21='m'*'dursq'
20   calc c22='c'*'agestart'
21   calc c23='m'*'agestart'

22   name c18 'c_dur'
23   name c19 'm_dur'
24   name c20 'c_dursq'
25   name c21 'm_dursq'
26   name c22 'c_age'
27   name c23 'm_age'

28   NOTE: Declare age variables as categorical and name categories
29   catn 1 'c_age' 1 'c_age<20' 2 'c_age20-24' 3 'c_age25-29' 4 'c_age30+'
30   catn 1 'm_age' 1 'm_age<20' 2 'm_age20-24' 3 'm_age25-29' 4 'm_age30+'

31   NOTE: Declare response as multinomial
32   NOTE: The response is placed in C24 (equal to SEP in the binary case)
33   NOTE: A response indicator is placed C25 (equal to 1 for binary)
34   NOTE: Further details of the multinomial command given in Exercise 4 on competing risks
35   catn 1 'sep' 0 'none' 1 '\sep'
36   mnom 0 'sep' c24 c25 0

37   NOTE: Declare identifiers for different levels (1 and 2 are the same for a binary response)
38   NOTE: In the multilevel model we only specify random effects at the woman level
39   iden 1 c25 2 'lev1id' 3 'woman'

40   NOTE: Divide response by ATRISK to obtain probability of each type of event
41   NOTE: within 6-month interval
42   NOTE: Declare as response
43   NOTE: Create binomial response (rate of separation per month)
```

```
44   calc c29=c24/'atrisk'
45   name c29 'y'
46   resp 'y'

47   NOTE: Declare ATRISK as the denominator
48   calc c30='atrisk'
49   name c30 'denom'

50   NOTE: Add explanatory variables
51   addt 'c'
52   addt 'c_dur'
53   addt 'c_dursq'
54   addt 'c_age'
55   addt 'm'
56   addt 'm_dur'
57   addt 'm_dursq'
58   addt 'm_age'
```