Solutions to selected exercises

Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata (3rd Edition).* College Station, TX: Stata Press.

Volume II: Categorical Responses, Counts, and Survival

Contents

Disclaimer

We have solved the exercises as well as we could but there may be better solutions and we may have made mistakes. We are grateful for any suggestions for improvement.

Please also check the errata at http://www.stata.com/bookstore/mlmus3.html for any errors in the wording of the exercises themselves.

10.3 Vaginal-bleeding data

1. *Produce an identifier variable for women, and reshape the data to long form, stacking the responses* `y1`–`y4` *into one variable and creating a new variable,* `occasion`*, taking the values 1–4 for each woman.*

```
. use amenorrhea, clear
. generate id = _n
. reshape long y, i(id) j(occasion)
(note: j = 1 2 3 4)
Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       57   ->     228
Number of variables                   7   ->       5
j variable (4 values)                     ->   occasion
xij variables:
                             y1 y2 ... y4   ->   y
-----------------------------------------------------------------------------
```

2. *Fit the following model considered by Fitzmaurice, Laird, and Ware (2011):*

$$\text{logit}\{\Pr(y_{ij}=1|x_j,t_{ij},\zeta_j)\} \;=\; \beta_1+\beta_2 t_{ij}+\beta_3 t_{ij}^2+\beta_4 x_j t_{ij}+\beta_5 x_j t_{ij}^2+\zeta_j$$

where $t_{ij} = 1,2,3,4$ *is the time interval and* x_j *is* `dose`*. It is assumed that* $\zeta_j \sim N(0,\psi)$*, and that* ζ_j *is independent across women and independent of* x_j *and* t_{ij}*. Use* `gllamm` *with the* `weight(wt)` *option to specify that* `wt2` *are level-2 weights.*

```
. generate time = occasion
. generate dose_time = dose*time
. generate time2 = time^2
. generate dose_time2 = dose*time2
. gllamm y time time2 dose_time dose_time2, i(id) family(binomial) link(logit)
> weight(wt) adapt
number of level 1 units = 3616
number of level 2 units = 1151
Condition Number = 61.916104
gllamm model
log likelihood = -1934.6777
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	1.129487	.2681452	4.21	0.000	.603932	1.655042
time2	-.0414252	.0548016	-0.76	0.450	-.1488344	.065984
dose_time	.5646349	.1925201	2.93	0.003	.1873024	.9419674
dose_time2	-.1096827	.0496279	-2.21	0.027	-.2069516	-.0124137
_cons	-3.796641	.3041371	-12.48	0.000	-4.392739	-3.200544

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id)
    var(1): 5.0152062 (.57035023)
------------------------------------------------------------------------------
```

3. *Write down the above model, but with a random slope of t_{ij} and fit the model.*

$$\text{logit}\{\Pr(y_{ij}=1|x_j,t_{ij},\zeta_j)\} \;=\; \beta_1+\beta_2 t_{ij}+\beta_3 t_{ij}^2+\beta_4 x_j t_{ij}+\beta_5 x_j t_{ij}^2+\zeta_{1j}+\zeta_{2j}t_{ij},$$

where ζ_{1j} and ζ_{2j} are a random intercept and random slope of time, and are assumed to have a bivariate normal distribution with zero means, variances ψ_1 and ψ_2 and correlation ρ.

```
. generate one = 1
. eq inter: one
. eq slope: time
. gllamm y time time2 dose_time dose_time2, i(id)
>    nrf(2) eqs(inter slope) f(binom) l(logit) weight(wt) adapt
number of level 1 units = 3616
number of level 2 units = 1151

Condition Number = 77.302239

gllamm model

log likelihood = -1927.1165
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	.8112482	.3460057	2.34	0.019	.1330896	1.489407
time2	.0184146	.0660331	0.28	0.780	-.1110078	.147837
dose_time	.5473806	.1973645	2.77	0.006	.1605533	.9342078
dose_time2	-.0987989	.0534429	-1.85	0.065	-.2035451	.0059473
_cons	-3.441387	.4534743	-7.59	0.000	-4.330181	-2.552594

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id)

    var(1): 4.6391244 (1.6963202)
    cov(2,1): -.34099882 (.42322007) cor(2,1): -.22015592

    var(2): .51714101 (.19987784)
------------------------------------------------------------------------------
```

4. *Interpret the estimated coefficients.*

The model assumes that there is no difference in the log-odds of amenorrhea between the groups at time 0 (baseline). In the low-dose group, the log-odds increase approximately by the same amount of 0.81 in each 3-month interval (since the estimated coefficient of `time2` is small and nonsignificant), corresponding to an odds ratio of 2.3. The interaction between `dose` and `time2` is not quite significant, so we could assume a linear relationship for both group by removing the terms `dose_time2` and `time2`. However, keeping the terms in, the high-dose group initially has a larger slope than the low-dose group, and the slope decreases over time because time-squared has a negative coefficient ($.0184 - .0988$).

5. *Plot marginal predicted probabilities as a function of time, separately for women in the two treatment groups.*

```
. gllapred prob, mu marg
(mu will be stored in prob)

. sort dose id time

. twoway (line prob time if dose==0, sort) (line prob time if dose==1, sort),
>  ytitle(Predicted marginal probability) xtitle(Time in 90 day intervals)
>  legend(order(1 "Low dose" 2 "High dose"))
```

The graph is shown in figure 1.

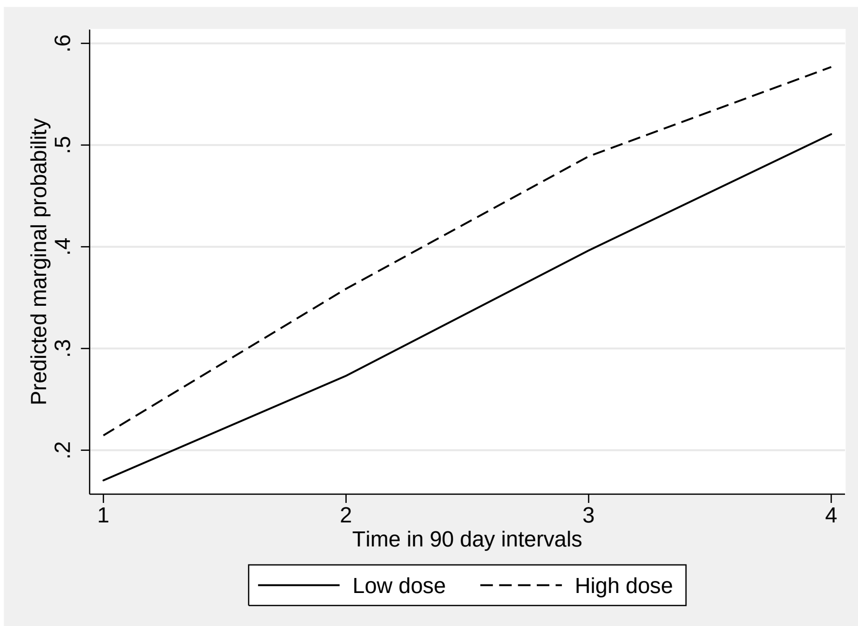

Figure 1: Predicted marginal probabilities over time by dose level

10.8 PISA data

1. *Fit a logistic regression model with* `pass_read` *as the response variable and the variables* `female` *to* `both_for` *above as covariates and with a random intercept for schools using* `gllamm`*. (Use the default eight quadrature points.)*

```
. use pisaUSA2000, clear

. gllamm pass_read female isei high_school college test_lang
>    one_for both_for, i(id_school) link(logit) family(binomial) adapt

number of level 1 units = 2069
number of level 2 units = 148

Condition Number = 335.04344

gllamm model

log likelihood = -1252.8108
```

pass_read	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	.5422157	.1031921	5.25	0.000	.3399629	.7444685
isei	.0206763	.003284	6.30	0.000	.0142397	.0271129
high_school	.4447949	.2565116	1.73	0.083	-.0579587	.9475484
college	.7968813	.2550522	3.12	0.002	.2969882	1.296774
test_lang	.7825116	.2834802	2.76	0.006	.2269005	1.338123
one_for	.0112568	.2244283	0.05	0.960	-.4286147	.4511283
both_for	.1507844	.2376408	0.63	0.526	-.314983	.6165517
_cons	-3.279322	.3811213	-8.60	0.000	-4.026306	-2.532339

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id_school)

    var(1): .51343023 (.12840606)
------------------------------------------------------------------------------
```

2. *Fit the model from step 1 with the school mean of* `isei` *as an additional covariate. (Use the estimates from step 1 as starting values.)*

```
. egen mn_isei = mean(isei), by(id_school)

. matrix a=e(b)
```

(*Continued on next page*)

```
. gllamm pass_read female isei mn_isei high_school college test_lang
>    one_for both_for, i(id_school) link(logit) family(binomial) from(a) adapt

number of level 1 units = 2069
number of level 2 units = 148

Condition Number = 595.81116

gllamm model

log likelihood = -1225.4697
```

pass_read	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	.5552102	.102912	5.39	0.000	.3535063	.7569141
isei	.0143423	.003335	4.30	0.000	.0078058	.0208787
mn_isei	.0690721	.0092476	7.47	0.000	.0509472	.0871971
high_school	.3999544	.2561423	1.56	0.118	-.1020752	.901984
college	.720787	.254843	2.83	0.005	.2213039	1.22027
test_lang	.6951882	.2849895	2.44	0.015	.1366191	1.253757
one_for	-.0199179	.2239413	-0.09	0.929	-.4588347	.418999
both_for	.0986698	.2359626	0.42	0.676	-.3638083	.561148
_cons	-6.033619	.5387262	-11.20	0.000	-7.089502	-4.977735

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id_school)

    var(1): .27143215 (.08570122)
------------------------------------------------------------------------------
```

3. *Interpret the estimated coefficients of* `isei` *and school mean* `isei` *and comment on the change in the other parameter estimates due to adding school mean* `isei`*.*

   Within a school, student's ISEI score has an estimated effect of 0.014 on the log-odds scale and between schools there is an additional effect of 0.069. Considering a 10-unit change in ISEI, the corresponding odds ratios are 1.15 ($= \exp(0.14)$) and 2.00 ($= \exp(0.69)$). Comparing two students from the same school, one of whom has ISEI 10 points higher than the other (with all other covariates being the same), the higher ISEI student has a 15% greater odds of passing the reading test. Comparing two students with the same ISEI score (and other covariate values) from schools that differ in their mean ISEI score by 10 units (but have the same random intercept), the student from the higher mean ISEI school has twice the odds of passing the reading test as the other student.

   The estimated random intercept variance has nearly halved due to adding school mean ISEI. The estimates of the effects of parent's education on test language spoken at home have decreased a little.

4. *From the estimates in step 2, obtain an estimate of the between-school effect of socioeconomic status.*

   The total between-school effect on the log-odds scale is the sum of the coefficient of `isei` and `mn_isei`, giving 0.083 ($= 0.014 + 0.069$).

5. *Obtain robust standard errors using the command* `gllamm, robust`, *and compare them with the model-based standard errors.*

```
. gllamm, robust
Non-adaptive log-likelihood: -1225.4744
-1225.4697 -1225.4697
number of level 1 units = 2069
number of level 2 units = 148

Condition Number = 595.81116

gllamm model

log likelihood = -1225.4697

Robust standard errors
```

pass_read	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	.5552102	.1024602	5.42	0.000	.3543919	.7560284
isei	.0143423	.0029873	4.80	0.000	.0084873	.0201972
mn_isei	.0690721	.0090417	7.64	0.000	.0513507	.0867935
high_school	.3999544	.2619124	1.53	0.127	-.1133844	.9132932
college	.720787	.2574594	2.80	0.005	.2161759	1.225398
test_lang	.6951882	.269443	2.58	0.010	.1670896	1.223287
one_for	-.0199179	.1998363	-0.10	0.921	-.4115898	.3717541
both_for	.0986698	.2452364	0.40	0.687	-.3819847	.5793244
_cons	-6.033619	.5471276	-11.03	0.000	-7.105969	-4.961268

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id_school)

    var(1): .27143215 (.08152135)
------------------------------------------------------------------------------
```

The robust and model-based standard errors are quite similar in this case.

(*Continued on next page*)

6. *Add a random coefficient of* `isei`*, and compare the random-intercept and random-coefficient models using a likelihood ratio test. Use the estimates from step 2 (or step 5) as starting values, adding zeros for the two additional parameters as shown in section 11.7.2.*

```
. estimates store ri
. generate one = 1
. eq inter: one
. eq slope: isei
. matrix a=e(b)
. matrix a = (a, 0, 0)
. gllamm pass_read female isei mn_isei high_school college test_lang
>    one_for both_for, i(id_school) link(logit) family(binomial) adapt
>    from(a) copy nrf(2) eqs(inter slope)
number of level 1 units = 2069
number of level 2 units = 148

Condition Number = 615.85825

gllamm model

log likelihood = -1225.1738
```

pass_read	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	.553834	.1028715	5.38	0.000	.3522096	.7554583
isei	.0147346	.0033942	4.34	0.000	.0080822	.021387
mn_isei	.0685597	.0092542	7.41	0.000	.0504219	.0866976
high_school	.4042809	.2573612	1.57	0.116	-.1001377	.9086994
college	.7320911	.2565077	2.85	0.004	.2293452	1.234837
test_lang	.69372	.2850084	2.43	0.015	.1351137	1.252326
one_for	-.0198951	.2240206	-0.09	0.929	-.4589673	.4191771
both_for	.0948089	.2357538	0.40	0.688	-.3672601	.5568779
_cons	-6.040813	.5397841	-11.19	0.000	-7.09877	-4.982855

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id_school)

    var(1): .45755695 (.29755562)
    cov(2,1): -.00214805 (.00344746) cor(2,1): -1

    var(2): .00001008 (.00002619)
------------------------------------------------------------------------------
```

We can already see that the random-slope variance estimate is close to zero and that the log likelihood has not changed much. The likelihood ratio test confirms that there is no evidence for a random slope:

```
. estimates store rc
. lrtest ri rc
Likelihood-ratio test                                 LR chi2(2)  =      0.59
(Assumption: ri nested in rc)                         Prob > chi2 =    0.7439
```

7. ❖ *In this survey, schools were sampled with unequal probabilities, π_j, and given that a school was sampled, students were sampled from the school with unequal probabilities $\pi_{i|j}$. The reciprocals of these probabilities are given as school- and student-level survey weights,* `wnrschbg` *($w_j = 1/\pi_j$) and* `w_fstuwt` *($w_{i|j} = 1/\pi_{i|j}$), respectively. As discussed in Rabe-Hesketh and Skrondal (2006), incorporating survey weights in multilevel models using a so-called pseudolikelihood approach can lead to biased estimates, particularly if the level-1 weights $w_{i|j}$ are very different from 1 and if the cluster sizes are small. Neither of these issues arise here, so implement pseudo maximum likelihood estimation as follows:*

   a. *Rescale the student-level weights by dividing them by their cluster means [this is scaling method 2 in Rabe-Hesketh and Skrondal (2006)].*

```
. egen mnw = mean(w_fstuwt), by(id_school)
. generate wt1 = w_fstuwt/mnw
```

   b. *Rename the level-2 weights and rescaled level-1 weights to* `wt2` *and* `wt1`*, respectively.*

```
. rename wnrschbw wt2
```

   c. *Run the* `gllamm` *command from step 2 above with the additional option* `pweight(wt)` *(Only the stub of the weight variables is specified;* `gllamm` *will look for the level-1 weights under* `wt1` *and the level-2 weights under* `wt2`*.) Use the estimates from step 2 as starting values.*

```
. matrix a=e(b)
. gllamm pass_read female isei mn_isei high_school college test_lang
>    one_for both_for, i(id_school) link(logit) family(binomial) from(a)
>    pweight(wt) adapt
number of level 1 units = 2069
number of level 2 units = 148
Condition Number = 634.97035
gllamm model
log likelihood = -197964.36
Robust standard errors
```

pass_read	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
female	.6218816	.1540693	4.04	0.000	.3199114	.9238518
isei	.0182009	.0048055	3.79	0.000	.0087824	.0276194
mn_isei	.0682412	.0164297	4.15	0.000	.0360395	.1004429
high_school	.1019586	.4766681	0.21	0.831	-.8322938	1.036211
college	.4528054	.5050717	0.90	0.370	-.537117	1.442728
test_lang	.6245943	.3825914	1.63	0.103	-.125271	1.37446
one_for	-.1086344	.2740453	-0.40	0.692	-.6457533	.4284845
both_for	-.2811828	.3265269	-0.86	0.389	-.9211638	.3587982
_cons	-5.875254	.95455	-6.15	0.000	-7.746138	-4.004371

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id_school)

    var(1): .29620737 (.12431098)
------------------------------------------------------------------------------
```

d. *Compare the estimates with those from step 2. Robust standard errors are computed by* `gllamm` *because model-based standard errors are not appropriate with survey weights.*

   Some of the estimates are quite different, especially the coefficients of `high_school` and `college`.

11.7 Recovery after surgery data

1. *Reshape the data to long form, stacking the recovery scores at the four occasions into a single variable and generating an identifier,* `occ`*, for the four occasions. (You can specify several variables in the* `i()` *option of the* `reshape` *command if one variable does not uniquely identify the individuals.) Recode the recovery score to four categories (to simplify some of the commands below), by merging {0,1}, {2,3}, and {4,5} and calling the new categories 1, 2, 3, and 4.*

```
. use recovery, clear
. reshape long score, i(id dosage) j(occ)
(note: j = 1 2 3 4)
Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                       60   ->     240
Number of variables                   8   ->       6
j variable (4 values)                     ->   occ
xij variables:
               score1 score2 ... score4   ->   score
-----------------------------------------------------------------------------
```

Before we forget, let us construct a unique person identifier

```
. egen id2 = group(id dosage)
```

Now recode the response variable:

```
. recode score 0/1=1 2/3=2 4/5=3 6=4
(score: 164 changes made)
```

2. *Construct a variable,* `time`*, taking the values 0, 5, 15, and 30 at the four occasions. Fit a random-intercept proportional odds model model with dummy variables for the dosage groups,* `age`*,* `duration`*, and* `time` *as covariates. (Make sure there are 60 level-2 clusters.)*

```
. recode occ 1=0 2=5 3=15 4=30, generate(time)
(240 differences between occ and time)
. tabulate dosage, generate(dose)

     dosage |      Freq.     Percent        Cum.
------------+-----------------------------------
         15 |         60       25.00       25.00
         20 |         60       25.00       50.00
         25 |         60       25.00       75.00
         30 |         60       25.00      100.00
------------+-----------------------------------
      Total |        240      100.00
```

(Continued on next page)

```
. gllamm score dose2 dose3 dose4 age duration time, i(id2)
>   link(ologit) adapt

number of level 1 units = 240
number of level 2 units = 60

Condition Number = 722.4517

gllamm model

log likelihood = -221.61016
```

score	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
score						
dose2	-.2008751	1.485427	-0.14	0.892	-3.112259	2.710509
dose3	-1.225982	1.430519	-0.86	0.391	-4.029748	1.577785
dose4	-1.801015	1.450042	-1.24	0.214	-4.643045	1.041015
age	-.0518457	.0345485	-1.50	0.133	-.1195595	.0158681
duration	-.022422	.0143755	-1.56	0.119	-.0505975	.0057536
time	.2352171	.0267231	8.80	0.000	.1828408	.2875934
_cut11						
_cons	-4.030454	2.091063	-1.93	0.054	-8.128863	.0679542
_cut12						
_cons	-1.255637	2.062151	-0.61	0.543	-5.297379	2.786105
_cut13						
_cons	1.449118	2.062102	0.70	0.482	-2.592527	5.490763

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id2)

    var(1): 13.396126 (4.1176567)
------------------------------------------------------------------------------
```

3. *Compare the model from step 2 with a model including* `dosage` *as a continuous covariate instead of the dummy variables for dosage groups, using a likelihood ratio test at the 5% significance level.*

```
. estimates store model1
```

```
. gllamm score dosage age duration time, i(id2) link(ologit) adapt

number of level 1 units = 240
number of level 2 units = 60

Condition Number = 932.73796

gllamm model

log likelihood = -221.66103
```

score	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
score						
dosage	-.1277787	.0920357	-1.39	0.165	-.3081654	.0526079
age	-.0553655	.0326618	-1.70	0.090	-.1193814	.0086505
duration	-.022134	.0142646	-1.55	0.121	-.050092	.0058241
time	.2350984	.0267134	8.80	0.000	.182741	.2874558
_cut11						
_cons	-6.208584	2.80452	-2.21	0.027	-11.70534	-.711825
_cut12						
_cons	-3.434832	2.759577	-1.24	0.213	-8.843504	1.97384
_cut13						
_cons	-.7321253	2.734417	-0.27	0.789	-6.091483	4.627233

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id2)

    var(1): 13.38398 (4.1172343)
------------------------------------------------------------------------------
. estimates store model2
. lrtest model1 .
Likelihood-ratio test                                 LR chi2(2)  =      0.10
(Assumption: model2 nested in model1)                 Prob > chi2 =    0.9504
```

Linearity of the log-odds for the covariate `dosage` is not rejected at the 5% level ($L = 0.10$, df $= 2$, $p = 0.95$).

4. *Extend the model chosen in step 3 to include an interaction between* `dosage` *and* `time`*. Test the interaction using a Wald test at the 5% level of significance.*

```
. matrix a=e(b)
. generate dosage_time = dosage*time
```

```
. gllamm score dosage age duration time dosage_time, i(id2) link(ologit)
>    adapt from(a)

number of level 1 units = 240
number of level 2 units = 60

Condition Number = 7708.0541

gllamm model

log likelihood = -221.48703
```

score	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
score						
dosage	-.1502809	.1007209	-1.49	0.136	-.3476902	.0471284
age	-.0551556	.0329607	-1.67	0.094	-.1197574	.0094461
duration	-.0223082	.0143929	-1.55	0.121	-.0505177	.0059014
time	.1985694	.0669927	2.96	0.003	.067266	.3298727
dosage_time	.0016908	.0028821	0.59	0.557	-.003958	.0073396
_cut11						
_cons	-6.698538	2.956525	-2.27	0.023	-12.49322	-.9038561
_cut12						
_cons	-3.919558	2.909429	-1.35	0.178	-9.621934	1.782817
_cut13						
_cons	-1.201271	2.876031	-0.42	0.676	-6.838188	4.435646

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id2)

    var(1): 13.649205 (4.2281563)
------------------------------------------------------------------------------
```

The dosage by time interaction is not significant at the 5% level ($z = 0.59$, $p = 0.56$).

(*Continued on next page*)

5. *For the model selected in step 4, interpret the estimated odds ratios and random-intercept variance.*

```
. estimates restore model2
. gllamm, eform

number of level 1 units = 240
number of level 2 units = 60

Condition Number = 932.73796

gllamm model

log likelihood = -221.66103
```

score	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
score						
dosage	.8800481	.0809958	-1.39	0.165	.7347938	1.054016
age	.9461393	.0309026	-1.70	0.090	.8874693	1.008688
duration	.9781092	.0139523	-1.55	0.121	.9511419	1.005841
time	1.265033	.0337934	8.80	0.000	1.200503	1.333032
_cut11						
_cons	-6.208584	2.80452	-2.21	0.027	-11.70534	-.711825
_cut12						
_cons	-3.434832	2.759577	-1.24	0.213	-8.843504	1.97384
_cut13						
_cons	-.7321253	2.734417	-0.27	0.789	-6.091483	4.627233

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id2)

    var(1): 13.38398 (4.1172343)
------------------------------------------------------------------------------
```

Each extra gram of anesthetic per kilogram of weight is associated with an estimated 12% reduction in the odds of having a recovery score above a given cut-point, after controlling for covariates. This translates to a 72% ($-72 = 100(0.8800481^{10} - 1)$) reduction in the odds for a 10grams/kilogram increase. Each extra month of age is associated with an estimated 5% decrease in the odds of a high recovery score after controlling for the other covariates. For a one-year increase in age, the odds are estimated to decrease by 49% ($-49 = 100(0.9461393^{12} - 1)$). Each extra minute of surgery reduces the estimated odds of a high recovery score by 2%, corresponding to a 35% decrease ($-35 = 100(0.9781092^{20} - 1)$) every 20 minutes. Finally, the estimated odds of a high recovery score increase over time after admission to the recovery room, by 27% per minute, after controlling for the other covariates.

The estimated random-intercept variance is large, giving an estimated residual intraclass correlation of the latent responses of 0.80 ($= 13.38398/(13.38398 + \pi^2/3)$).

6. ❖ *Extend the model selected in step 4 by relaxing the proportional odds assumption for dosage (see section 11.2 on using the* `thresh()` *option in* `gllamm` *to relax proportional odds). Test whether the odds are proportional using a likelihood ratio test.*

```
. eq thr: dosage
. matrix a=e(b)
. gllamm score age duration time, i(id2)
>    link(ologit) thresh(thr) from(a) skip adapt

number of level 1 units = 240
number of level 2 units = 60

Condition Number = 920.15769

gllamm model

log likelihood = -217.92407
```

score	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
score						
age	-.059168	.0332891	-1.78	0.076	-.1244134	.0060774
duration	-.0221955	.0144873	-1.53	0.126	-.05059	.0061991
time	.2428607	.028066	8.65	0.000	.1878523	.2978691
_cut11						
dosage	.1970775	.1005708	1.96	0.050	-.0000376	.3941926
_cons	-7.890397	3.004708	-2.63	0.009	-13.77952	-2.001279
_cut12						
dosage	.0501135	.0972566	0.52	0.606	-.140506	.240733
_cons	-1.731605	2.866093	-0.60	0.546	-7.349045	3.885835
_cut13						
dosage	.13174	.1013879	1.30	0.194	-.0669768	.3304567
_cons	-.7971798	2.892524	-0.28	0.783	-6.466422	4.872063

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id2)

    var(1): 13.833558 (4.3011138)
------------------------------------------------------------------------------

. estimates store model3
. lrtest model2 model3
Likelihood-ratio test                                 LR chi2(2)  =      7.47
(Assumption: model2 nested in model3)                 Prob > chi2 =    0.0238
```

We reject the proportional odds assumption for dosage group at the 5% level ($L = 7.47$, df $= 2$, $p = 0.02$).

7. *For age equal to 37 months, duration equal to 80 minutes, and time in recovery room equal to 15 minutes, produce a graph of predicted marginal probabilities similar to figure 11.13 for the model selected in step 6 or for the model selected in step 4. Also produce a stacked bar chart, treating dosage group as categorical.*

   First we set the explanatory variables equal to the required values and restore the estimates for model 2:

```
. replace age=37
(232 real changes made)
. replace duration=80
(240 real changes made)
. replace time=15
(180 real changes made)
. estimates restore model2
```

   Now we can predict the marginal probabilities using `gllamm`

```
. gllapred pr1, marg mu above(1) fsample
(mu will be stored in pr1)
. gllapred pr2, marg mu above(2) fsample
(mu will be stored in pr2)
. gllapred pr3, marg mu above(3) fsample
```

   For the figure resembling figure 11.12, we need the cumulative probabilities that y is anything from 1 up to category s, for $s = 1, 2, 3, 4$

```
. generate pr12 = 1-pr2
. generate pr123 = 1-pr3
. generate pr1234 = 1
. twoway (area pr1 dosage, sort fintensity(inten10))
>    (rarea pr12 pr1 dosage, sort fintensity(inten50))
>    (rarea pr123 pr12 dosage, sort fintensity(inten70))
>    (rarea pr1234 pr123 dosage, sort fintensity(inten90)),
>    legend(order(1 "Prob(y=1)" 2 "Prob(y=2)" 3 "Prob(y=3)" 4 "Prob(y=4)"))
>         xtitle("dosage")
```

   The graphs are given in figure 2 for models 2 and 3 (for model 3, run all the above commands after restoring model 3).

   Note that the boundaries on the graph are not exactly parallel when the proportional odds assumption is made, but the logit transformation of the boundaries is.

   For the bar chart, we need the probabilities that y equals each of the categories

```
. generate pr1is = 1-pr1
. generate pr2is = pr1 - pr2
. generate pr3is = pr2 - pr3
. generate pr4is = pr3
. graph bar (mean) pr1is pr2is pr3is pr4is, over(dosage) stack
> legend(order(1 "Pr(1)" 2 "Pr(1)" 3 "Pr(1)" 4 "Pr(1)"))
```

   The graphs are given in figure 3 for models 2 and 3 (for model 3, run all the above commands after restoring model 3).

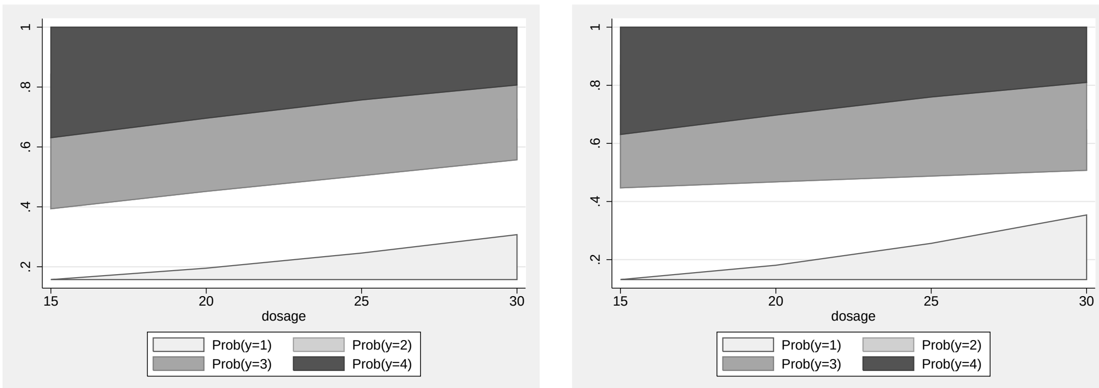

Figure 2: Area graphs of predicted marginal probabilities versus dosage groups, when age is 37 months, duration of surgery is 80 minutes, and recovery time is 15 minutes. Left panel is proportional odds model (model 2) and right panel relaxes proportional odds for dosage (model 3)

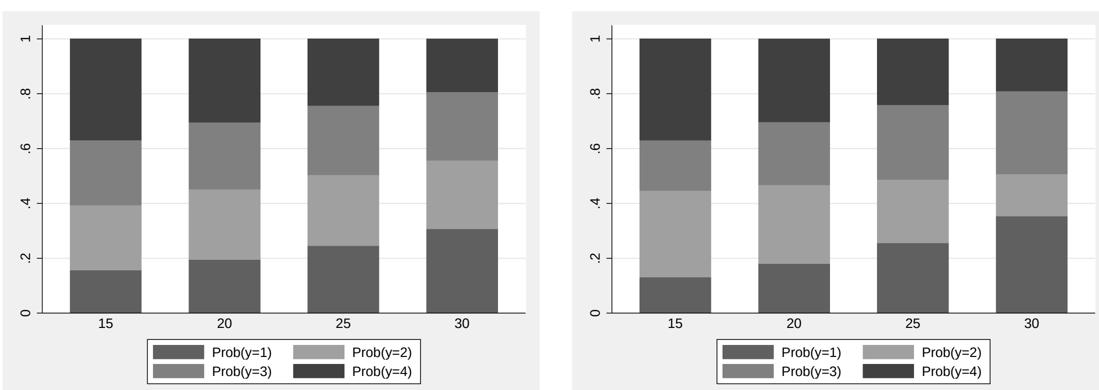

Figure 3: Stacked bar chart of predicted marginal probabilities for the dosage groups, when age is 37 months, duration of surgery is 80 minutes, and recovery time is 15 minutes. Left panel is proportional odds model (model 2) and right panel relaxes proportional odds for dosage (model 3)

12.4 British election data

1. *Create a variable,* `chosen`*, equal to 1 for the party voted for (*`rank` *equal to 1) and 0 for the other parties.*

```
. use elections, clear
. generate chosen = rank == 1
```

2. *Standardize* `lrdist` *and* `inflation` *to have mean 0 and variance 1. Produce all the dummy variables and interactions necessary to fit a conditional logistic regression model (using* `clogit`*) for* `chosen`*, with the following covariates: the standardized versions of* `lrdist` *and* `inflation`*, and the dummy varibles* `yr87`*,* `yr92`*,* `male`*, and* `manual`*. All variables except the standardized version of* `lrdist` *should have party-specific coefficients.*

```
. egen inflat = std(inflation)
. egen dist = std(rldist)
. tabulate party, generate(p)
      party |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      2,458       33.33       33.33
          2 |      2,458       33.33       66.67
          3 |      2,458       33.33      100.00
------------+-----------------------------------
      Total |      7,374      100.00
. rename p1 cons
. rename p2 lab
. rename p3 lib
. foreach var of varlist male inflat manual yr87 yr92 {
  2.         generate lab_`var' = lab*`var'
  3.         generate lib_`var' = lib*`var'
  4. }
```

3. *Fit the model using* `clogit` *and* `gllamm`*, using Conservatives as the base outcome.*

```
. clogit chosen dist lab_*  lib_* , group(occ)
Conditional (fixed-effects) logistic regression   Number of obs   =       7374
                                                  LR chi2(11)     =    1434.69
                                                  Prob > chi2     =     0.0000
Log likelihood = -1983.0429                       Pseudo R2       =     0.2656

------------------------------------------------------------------------------
      chosen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dist |  -1.134582   .0463711   -24.47   0.000    -1.225468   -1.043696
    lab_male |  -.7170468   .1247135    -5.75   0.000    -.9614808   -.4726129
  lab_inflat |     .40281   .0665768     6.05   0.000     .2723219     .533298
  lab_manual |   .5855308   .1298537     4.51   0.000     .3310223    .8400393
    lab_yr87 |  -.9940042   .1434858    -6.93   0.000    -1.275231   -.7127771
    lab_yr92 |  -.9786174   .1346003    -7.27   0.000    -1.242429   -.7148056
    lib_male |  -.6562548   .1194879    -5.49   0.000    -.8904468   -.4220627
  lib_inflat |   .3102374   .0623362     4.98   0.000     .1880607    .4324142
  lib_manual |  -.1422657   .1191864    -1.19   0.233    -.3758667    .0913353
    lib_yr87 |   -.785426   .1258898    -6.24   0.000    -1.032166   -.5386865
    lib_yr92 |  -1.068714   .1228379    -8.70   0.000    -1.309472   -.8279564
------------------------------------------------------------------------------
```

```
. gllamm party dist lab_* lib_*, nocons i(occ) link(mlogit)
>  expanded(occ chosen o) init

number of level 1 units = 7374

Condition Number = 7.2688994

gllamm model

log likelihood = -1983.0429
```

party	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dist	-1.134582	.0463711	-24.47	0.000	-1.225468	-1.043696
lab_male	-.7170468	.1247135	-5.75	0.000	-.9614807	-.4726128
lab_inflat	.40281	.0665768	6.05	0.000	.272322	.5332981
lab_manual	.585531	.1298537	4.51	0.000	.3310225	.8400395
lab_yr87	-.9940045	.1434858	-6.93	0.000	-1.275232	-.7127774
lab_yr92	-.9786177	.1346003	-7.27	0.000	-1.24243	-.7148059
lib_male	-.6562546	.1194879	-5.49	0.000	-.8904466	-.4220626
lib_inflat	.3102375	.0623362	4.98	0.000	.1880608	.4324142
lib_manual	-.1422654	.1191864	-1.19	0.233	-.3758663	.0913356
lib_yr87	-.7854264	.1258898	-6.24	0.000	-1.032166	-.5386869
lib_yr92	-1.068715	.1228379	-8.70	0.000	-1.309473	-.8279569

4. *Extend the model to include a person-level random slope for* `lrdist`, *and fit the extended model in* `gllamm`.

```
. eq slope: dist
. gllamm party dist lab_* lib_*, nocons i(serialno) eqs(slope)
>    link(mlogit) expanded(occ chosen o) adapt
number of level 1 units = 7374
number of level 2 units = 1344
Condition Number = 8.1833746
gllamm model
log likelihood = -1940.6814
```

party	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dist	-1.668452	.0940924	-17.73	0.000	-1.85287	-1.484034
lab_male	-.8026911	.1458909	-5.50	0.000	-1.088632	-.5167502
lab_inflat	.4823476	.0791058	6.10	0.000	.3273031	.6373922
lab_manual	.6978195	.1536384	4.54	0.000	.3966939	.9989452
lab_yr87	-1.088198	.1658249	-6.56	0.000	-1.413209	-.763187
lab_yr92	-1.11707	.1563765	-7.14	0.000	-1.423562	-.8105775
lib_male	-.720465	.1354692	-5.32	0.000	-.9859797	-.4549503
lib_inflat	.3920127	.0718925	5.45	0.000	.251106	.5329194
lib_manual	-.0866056	.136415	-0.63	0.526	-.3539742	.180763
lib_yr87	-.8391223	.1426198	-5.88	0.000	-1.118652	-.5595927
lib_yr92	-1.177754	.1386755	-8.49	0.000	-1.449553	-.9059552

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (serialno)

    var(1): 1.0384731 (.19574625)
------------------------------------------------------------------------------
```

5. *Write down the model and interpret the estimates.*

The following model is specified for the conditional probability that party s is chosen by respondent j at occasion i, given the covariates and the random coefficient ζ_{2j} for `lrdist`:

$$\Pr(y_{ij}=s|x_{2ij}^{[s]},\mathbf{x}_{ij},\zeta_{2j})$$
$$= \frac{\exp\left\{(\beta_2+\zeta_{2j})x_{2ij}^{[s]} + \beta_3^{[s]}x_{3j} + \beta_4^{[s]}x_{4ij} + \beta_5^{[s]}x_{5j} + \beta_6^{[s]}x_{6i} + \beta_7^{[s]}x_{7i}\right\}}{\sum_{c=1}^{3}\exp\left\{(\beta_2+\zeta_{2j})x_{2ij}^{[c]} + \beta_3^{[c]}x_{3j} + \beta_4^{[c]}x_{4ij} + \beta_5^{[c]}x_{5j} + \beta_6^{[c]}x_{6i} + \beta_7^{[c]}x_{7i}\right\}}$$

Here $x_{2ij}^{[s]}$ represents `lrdist` for party s, x_{3j} represents `male`, x_{4ij} represents `inflation`, x_{5j} represents `manual`, x_{6i} represents `yr87`, and x_{7i} represents `yr92`. It is assumed that the random coefficient ζ_{2j} has a normal distribution with zero mean and variance ψ, and that the covariates are independent of the random coefficient.

We now turn to the interpretation of the estimates. Controlling for the other covariates, the conditional or respondent-specific odds of choosing a party decreases by 81% (-81% = 100% $\times$ $\exp(-1.668452)-1$) as the distance between the party and the respondent on the left-right political dimension increases by one unit. The variance of the respondent-specific effects $\beta_2+\zeta_{2j}$ is estimated as 1.0384731 so a 95% range of the odds ratio is $(\exp(-1.668452-1.96\sqrt{1.0384731}, \exp(-1.668452-1.96\sqrt{1.0384731}) = (0.03, 1.39)$.

The following interpretations are all in terms of conditional odds with Conservatives as base-category and given the other covariates.

We first consider the odds of choosing Labour. The odds of choosing Labour in 1987 is estimated as 0.34=$\exp(-1.088198)$ when all covariates are zero. The odds of choosing Labour in 1992 is estimated as 0.33=$\exp(-1.11707)$ when all covariates are zero. The odds of choosing Labour is estimated as 55% (-55% = 100% $(\exp(-0.8026911)-1)$) lower for males than for females. The odds of choosing Labour is estimated as 62% (62% = 100% $(\exp(0.4823476)-1)$) higher when the perceived inflation rating increases by one unit (which might be explained by the fact that Conservatives were the incumbents). The odds of choosing Labour is estimated as 100% (100% = 100% $(\exp(0.6978195)-1)$) higher for respondents whose father was a manual worker compared to the father not being a manual worker.

We then consider the odds of choosing Liberals. The odds of choosing Liberals in 1987 is estimated as 0.43=$\exp(-0.8391223)$ when all covariates are zero. The odds of choosing Liberals in 1992 is estimated as 0.31=$\exp(-1.177754)$ when all covariates are zero. The odds of choosing Liberals is estimated as 51% (-51% = 100% $(\exp(-0.720465)-1)$) lower for males than for females. The odds of choosing Liberals is estimated as 34% (34% = 100% $(\exp(0.2920127)-1)$) higher when the perceived inflation rating increases by one unit (which might be explained by the fact that Conservatives were the incumbents). The odds of choosing Liberals is estimated as 8% (-8% = 100% $(\exp(-0.0866056)-1)$) lower for respondents whose father was a manual worker compared to the father not being a manual worker.

(*Continued on next page*)

6. *Instead of including a random slope for* `lrdist`*, include correlated person-level random intercepts for Labour and Liberal. Use the options* `ip(m)` *and* `nip(15)` *to use degree-15 spherical quadrature. This problem will take quite a long time to run.*

```
. gllamm party dist lab_* lib_*, nocons i(serialno) nrf(2) eqs(lab lib)
>    link(mlogit) expanded(occ chosen o) ip(m) nip(15) trace adapt

number of level 1 units = 7374
number of level 2 units = 1344

Condition Number = 10.914979

gllamm model

log likelihood = -1789.9395
```

party	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dist	-2.040746	.132729	-15.38	0.000	-2.30089	-1.780602
lab_male	-1.223027	.3104283	-3.94	0.000	-1.831455	-.6145988
lab_inflat	.7402028	.1378139	5.37	0.000	.4700924	1.010313
lab_manual	1.439134	.3339982	4.31	0.000	.7845092	2.093758
lab_yr87	-1.969602	.354374	-5.56	0.000	-2.664163	-1.275042
lab_yr92	-1.821407	.3324202	-5.48	0.000	-2.472938	-1.169875
lib_male	-1.108516	.3069595	-3.61	0.000	-1.710146	-.5068865
lib_inflat	.6237005	.13012	4.79	0.000	.36867	.8787311
lib_manual	.0918619	.3194567	0.29	0.774	-.5342618	.7179856
lib_yr87	-1.58963	.3265276	-4.87	0.000	-2.229612	-.9496478
lib_yr92	-2.054298	.329796	-6.23	0.000	-2.700687	-1.40791

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (serialno)

    var(1): 12.496053 (2.2906096)
    cov(2,1): 9.8049199 (1.8391321) cor(2,1): .77256248

    var(2): 12.889854 (1.9735669)
------------------------------------------------------------------------------
```

13.1 Epileptic-fit data

1. *Model II in Breslow and Clayton is a log-linear (Poisson regression) model with covariates* `lbas`, `treat`, `lbas_trt`, `lage`, *and* `v4`, *and a normally distributed random intercept for subjects. Fit this model using* `gllamm`.

```
. use epilep, clear
. gllamm y lbas treat lbas_trt lage v4, i(subj) link(log) family(poisson) adapt
number of level 1 units = 236
number of level 2 units = 59

Condition Number = 9.3178452

gllamm model

log likelihood = -665.29073
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lbas	.8844373	.1312305	6.74	0.000	.6272302	1.141644
treat	-.933037	.4008289	-2.33	0.020	-1.718647	-.1474268
lbas_trt	.3382596	.2033363	1.66	0.096	-.0602722	.7367914
lage	.4842389	.3472751	1.39	0.163	-.1964078	1.164886
v4	-.1610871	.0545758	-2.95	0.003	-.2680537	-.0541206
_cons	2.114294	.2197154	9.62	0.000	1.68366	2.544929

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (subj)

    var(1): .25282428 (.05894094)
------------------------------------------------------------------------------
```

(*Continued on next page*)

2. *Breslow and Clayton also considered a random-coefficient model (Model IV) using the variable* `visit` *instead of* `v4`. *The effect of* `visit` z_{ij} *varies randomly between subjects. The model can be written as*

$$\log(\mu_{ij}) \;=\; \beta_1 + \beta_2 x_{2j} + \cdots + \beta_5 x_{5j} + \beta_6 z_{ij} + \zeta_{1j} + \zeta_{2j} z_{ij}$$

where the subject-specific random intercept ζ_{1j} *and slope* ζ_{2j} *have a bivariate normal distribution, given the covariates. Fit this model using* `gllamm`.

```
. eq int: cons

. eq slope: visit

. gllamm y lbas treat lbas_trt lage visit, i(subj) link(log) family(poisson)
> nrf(2) eqs(int slope) ip(m) nip(15) adapt

number of level 1 units = 236
number of level 2 units = 59

Condition Number = 9.3163303

gllamm model

log likelihood = -655.681
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lbas	.8849772	.1312511	6.74	0.000	.6277298	1.142225
treat	-.9286564	.4021601	-2.31	0.021	-1.716876	-.1404371
lbas_trt	.3379746	.2044432	1.65	0.098	-.0627267	.7386759
lage	.4767191	.3536189	1.35	0.178	-.2163612	1.169799
visit	-.2664097	.1647101	-1.62	0.106	-.5892357	.0564162
_cons	2.09955	.2203692	9.53	0.000	1.667635	2.531466

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (subj)

    var(1): .25149333 (.05878604)
    cov(2,1): .00287153 (.08870133) cor(2,1): .00785428

    var(2): .53148135 (.2293816)
------------------------------------------------------------------------------
```

3. *Plot the posterior mean counts versus time for twelve patients in each treatment group.*

```
. gllapred pred, mu
(mu will be stored in pred)

. sort treat subj

. by treat subj: generate f=_n==1

. by treat: generate id=sum(f)

. twoway line pred visit if id<13 & treat==0, by(id)

. twoway line pred visit if id<13 & treat==1, by(id)
```

The graphs are shown in figures 4 and 5.

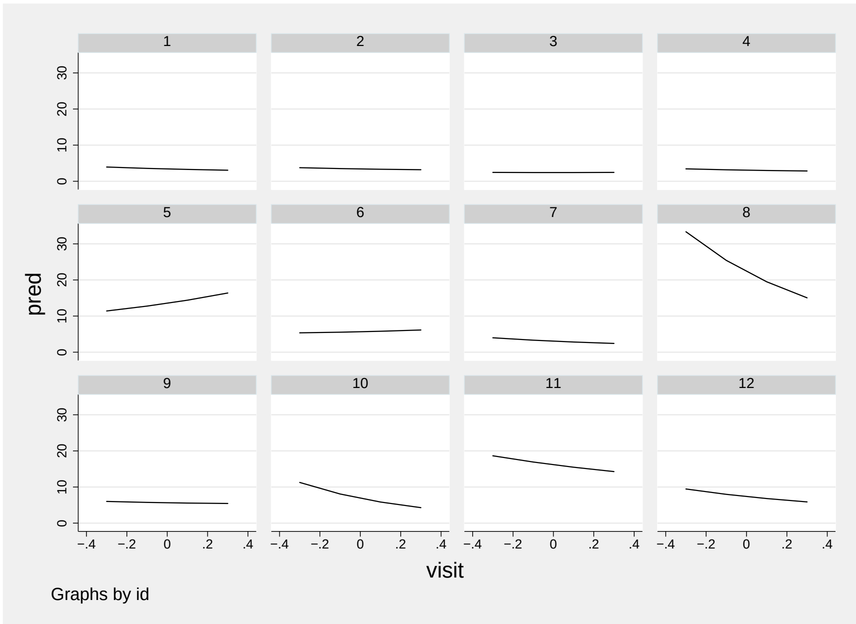

Figure 4: Posterior mean number of epileptic fits versus time for placebo group

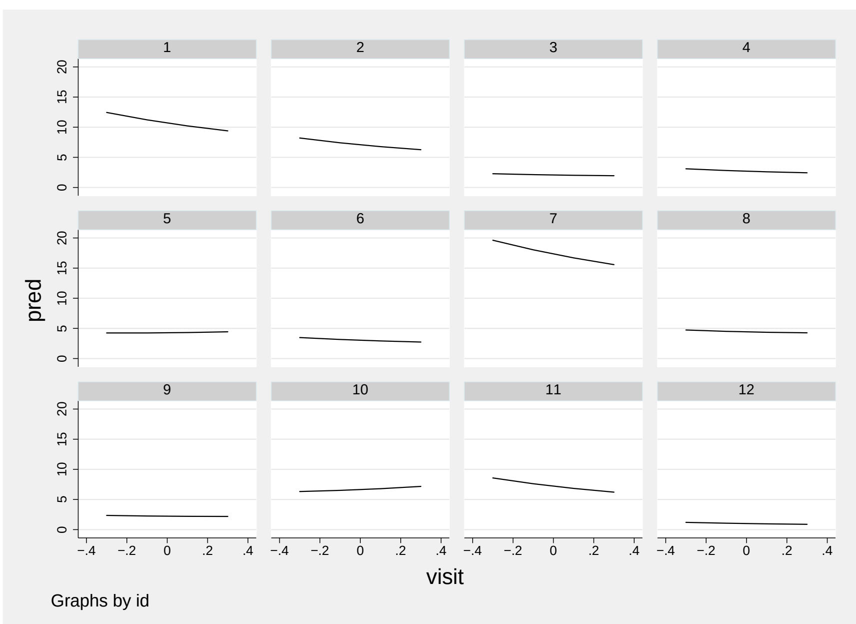

Figure 5: Posterior mean number of epileptic fits versus time for treatment group

14.7 Cigarette data

1. *Expand the data to person–period data.*

```
. use cigarette, clear

. generate id=_n

. expand time
(1670 observations created)

. by id, sort: gen t = _n

. generate y=0

. by id (t), sort: replace y = event if _n==_N
(634 real changes made)
```

2. *Estimate the discrete-time model that assumes the continuous-time hazards to be proportional. Include* `cc`, `tv`, *and their interaction as explanatory variables and specify a random intercept for classes. Use dummy variables for periods.*

```
. tabulate t, generate(occ)
          t |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      1,556       48.23       48.23
          2 |      1,082       33.54       81.77
          3 |        588       18.23      100.00
------------+-----------------------------------
      Total |      3,226      100.00

. xtset class
       panel variable:  class (unbalanced)

. xtcloglog y male cc tv cc_tv occ2 occ3
Random-effects complementary log-log model      Number of obs      =      3226
Group variable: class                           Number of groups   =       134

Random effects u_i ~ Gaussian                   Obs per group: min =         3
                                                               avg =      24.1
                                                               max =        54

                                                Wald chi2(6)       =     12.09
Log likelihood  = -1592.3537                    Prob > chi2        =    0.0599

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .0594819   .0804729     0.74   0.460    -.0982421    .2172059
          cc |   .1293571   .1216005     1.06   0.287    -.1089755    .3676896
          tv |   .0914655    .122232     0.75   0.454    -.1481048    .3310357
       cc_tv |  -.1605053   .1747717    -0.92   0.358    -.5030516    .1820409
        occ2 |   .0462722   .0918315     0.50   0.614    -.1337142    .2262586
        occ3 |   .3248201   .1042103     3.12   0.002     .1205717    .5290685
       _cons |  -1.707058   .1068043   -15.98   0.000     -1.91639   -1.497725
-------------+----------------------------------------------------------------
    /lnsig2u |  -3.357634   .8635395                      -5.05014   -1.665128
-------------+----------------------------------------------------------------
     sigma_u |   .1865946   .0805659                      .0800527    .4349328
         rho |   .0207278   .0175283                      .0038807    .1031386
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =     1.76 Prob >= chibar2 = 0.092
```

(*Continued on next page*)

3. *Interpret the exponentials of the estimated regression coefficients.*

```
. xtcloglog, eform

Random-effects complementary log-log model      Number of obs      =      3226
Group variable: class                           Number of groups   =       134

Random effects u_i ~ Gaussian                   Obs per group: min =         3
                                                               avg =      24.1
                                                               max =        54

                                                Wald chi2(6)       =     12.09
Log likelihood  = -1592.3537                    Prob > chi2        =    0.0599

------------------------------------------------------------------------------
           y |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   1.061287   .0854048     0.74   0.460     .9064294      1.2426
          cc |   1.138096   .1383931     1.06   0.287     .8967524    1.444394
          tv |   1.095779   .1339392     0.75   0.454     .8623408    1.392409
       cc_tv |   .8517133   .1488554    -0.92   0.358     .6046826    1.199663
        occ2 |   1.047359   .0961806     0.50   0.614       .87484      1.2539
        occ3 |   1.383782   .1442043     3.12   0.002     1.128142    1.697351
-------------+----------------------------------------------------------------
    /lnsig2u |  -3.357634   .8635395                      -5.05014   -1.665128
-------------+----------------------------------------------------------------
     sigma_u |   .1865946   .0805659                      .0800527    .4349328
         rho |   .0207278   .0175283                      .0038807    .1031386
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =     1.76 Prob >= chibar2 = 0.092
```

At the 5% level of significance there is not sufficient evidence to conclude that the interventions had any effects.

Specifically, for each intervention on its own (when the other intervention is not used), the hazard ratio does not differ significantly from 1. When combined with the other intervention, the hazard ratio for each intervention decreases by an estimated 15% (since the hazard ratio for the interaction is 0.85).

The hazards of smoking are estimated as 38% greater in 9th grade than in 7th grade after controlling for the other variables.

4. *Obtain the estimated residual intraclass correlation of the latent responses.*

This is given in the output under `rho` as 0.02. If you used `gllamm` to estimate the model, you can calculate the estimated intraclass correlation using

```
. display .1865946^2/(.1865946^2+_pi^2/6)
.02072779
```

This is a very small correlation, and we also see from the last line of the `xtcloglog` output that we cannot reject the null hypothesis (at the 5% level) that the true intraclass correlation is 0.

15.4 Bladder cancer data

1. *Wei, Lin, and Weissfeld (1989) specify a marginal Cox regression model based on total time and semi-restricted risk sets, where the risk set for a kth event includes risk intervals for all previous events ($< k$). They specify event-specific baseline hazards and allow the effects of* `treat`, `number`, *and* `size` *to differ between events. Fit this model.*

```
. use bladder, clear

. egen obs = group(enum id)

. stset stop, failure(event=1) id(obs)

                id:  obs
     failure event:  event == 1
obs. time interval:  (stop[_n-1], stop]
 exit on or before:  failure

------------------------------------------------------------------------------
      340  total obs.
        0  exclusions
------------------------------------------------------------------------------
      340  obs. remaining, representing
      340  subjects
      112  failures in single failure-per-subject data
     8522  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        59

. sort id enum

. list id enum start stop event _t0 _t _d _st if id>6&id<10 & _st==1, sepby(id)
```

	id	enum	start	stop	event	_t0	_t	_d	_st
25.	7	1	0	18	0	0	18	0	1
26.	7	2	18	18	0	0	18	0	1
27.	7	3	18	18	0	0	18	0	1
28.	7	4	18	18	0	0	18	0	1
29.	8	1	0	5	1	0	5	1	1
30.	8	2	5	18	0	0	18	0	1
31.	8	3	18	18	0	0	18	0	1
32.	8	4	18	18	0	0	18	0	1
33.	9	1	0	12	1	0	12	1	1
34.	9	2	12	16	1	0	16	1	1
35.	9	3	16	18	0	0	18	0	1
36.	9	4	18	18	0	0	18	0	1

The model could be parameterized by having a coefficient for `treat`, `number`, and `size`, as well as coefficients for interactions of each of these variables with dummy variables for the second, third and fourth events. Instead, we will include interactions between dummy variables for *each event*, including the first, and `treat`, `number`, and `size`. We must then omit "main effects" for `treat`, `number`, and `size`:

(*Continued on next page*)

```
. stcox ibn.enum#(c.treat c.number c.size), strata(enum) vce(cluster id) efron

         failure _d:  event == 1
   analysis time _t:  stop
                 id:  obs

Stratified Cox regr. -- Efron method for ties

No. of subjects      =          340             Number of obs   =        340
No. of failures      =          112
Time at risk         =         8522
                                                Wald chi2(12)   =      34.32
Log pseudolikelihood =   -423.73286             Prob > chi2     =     0.0006

                                    (Std. Err. adjusted for 85 clusters in id)

                         Robust
          _t  Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]

enum#c.treat
           1    .5909733   .1874038    -1.66   0.097     .3174264    1.100253
           2    .5313625   .1968685    -1.71   0.088     .2570531    1.098396
           3    .4973349   .2103116    -1.65   0.099     .2171177    1.139207
           4    .5297029   .2649767    -1.27   0.204     .1987149    1.411999

       enum#
    c.number
           1    1.268937   .0952058     3.17   0.002     1.095409    1.469955
           2    1.146744   .1012115     1.55   0.121     .9645825    1.363306
           3     1.18947   .1264058     1.63   0.103     .9658189    1.464911
           4    1.394411   .1621041     2.86   0.004      1.11029    1.751238

 enum#c.size
           1    1.072094   .0955849     0.78   0.435      .900206    1.276802
           2    .9251941   .1106043    -0.65   0.515     .7319378    1.169477
           3    .8074792   .1409972    -1.22   0.221     .5734553    1.137007
           4    .8134582   .1585875    -1.06   0.290     .5551233    1.192013

                                                             Stratified by enum
```

2. *Use* `testparm` *to test whether the coefficients of* `treat` *differ significantly between events (at the 5% level) and similarly for* `number` *and* `size`.

   In order to use `testparm`, it is better to use the more standard way of including interactions, where the dummy variable for event 1 is excluded and `treat`, `number`, and `size` are included:

(*Continued on next page*)

```
. stcox i.enum#(c.treat c.number c.size) c.treat c.number c.size,
>    strata(enum) vce(cluster id) efron

         failure _d:  event == 1
   analysis time _t:  stop
                 id:  obs

Stratified Cox regr. -- Efron method for ties

No. of subjects      =          340             Number of obs   =        340
No. of failures      =          112
Time at risk         =         8522
                                                Wald chi2(12)   =      34.32
Log pseudolikelihood =   -423.73286             Prob > chi2     =     0.0006

                                  (Std. Err. adjusted for 85 clusters in id)
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
enum#c.treat |
          2  |    .899131   .3020539    -0.32   0.752     .4654473    1.736903
          3  |   .8415522    .337499    -0.43   0.667     .3834531    1.846928
          4  |   .8963229   .4565585    -0.21   0.830     .3302854    2.432426
             |
       enum# |
    c.number |
          2  |   .9037042   .1068984    -0.86   0.392     .7167016      1.1395
          3  |   .9373751     .11348    -0.53   0.593     .7393767    1.188396
          4  |   1.098881   .1323528     0.78   0.434     .8678191    1.391464
             |
 enum#c.size |
          2  |   .8629789   .0990377    -1.28   0.199     .6891505    1.080653
          3  |   .7531798   .1153141    -1.85   0.064     .5579266    1.016764
          4  |   .7587567   .1442884    -1.45   0.147     .5226783    1.101465
             |
       treat |   .5909733   .1874038    -1.66   0.097     .3174264    1.100253
      number |   1.268937   .0952058     3.17   0.002     1.095409    1.469955
        size |   1.072094   .0955849     0.78   0.435      .900206    1.276802
------------------------------------------------------------------------------
                                                             Stratified by enum

. testparm enum#c.treat

 ( 1)  2.enum#c.treat = 0
 ( 2)  3.enum#c.treat = 0
 ( 3)  4.enum#c.treat = 0

           chi2(  3) =    0.24
         Prob > chi2 =    0.9715

. testparm enum#c.number

 ( 1)  2.enum#c.number = 0
 ( 2)  3.enum#c.number = 0
 ( 3)  4.enum#c.number = 0

           chi2(  3) =    5.86
         Prob > chi2 =    0.1186

. testparm enum#c.size

 ( 1)  2.enum#c.size = 0
 ( 2)  3.enum#c.size = 0
 ( 3)  4.enum#c.size = 0

           chi2(  3) =    3.61
         Prob > chi2 =    0.3065
```

None of the interactions are significant at the 5% level

3. *Fit the model by Wei, Lin, and Weissfeld (1989) but constraining all coefficients to be the same across events.*

```
. stcox treat number size, strata(enum) vce(cluster id) efron

         failure _d:  event == 1
   analysis time _t:  stop
                 id:  obs

Stratified Cox regr. -- Efron method for ties

No. of subjects      =          340              Number of obs   =        340
No. of failures      =          112
Time at risk         =         8522
                                                 Wald chi2(3)    =      15.35
Log pseudolikelihood =   -426.14683              Prob > chi2     =     0.0015

                                   (Std. Err. adjusted for 85 clusters in id)

             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .5572209   .1726125    -1.89   0.059     .3036319    1.022604
      number |    1.23404   .0827266     3.14   0.002       1.0821    1.407316
        size |   .9496925   .0903613    -0.54   0.587      .788121    1.144388
------------------------------------------------------------------------------
                                                             Stratified by enum
```

4. *In their model (2), Prentice, Williams, and Peterson (1981) use counting process risk intervals with restricted risk sets and event-specific baseline hazards. Fit this model, assuming that* `treat`, `number`, *and* `size` *have the same coefficients across events.*

```
. stset stop, enter(start) failure(event=1) id(obs)

                id:  obs
     failure event:  event == 1
obs. time interval:  (stop[_n-1], stop]
 enter on or after:  time start
 exit on or before:  failure

------------------------------------------------------------------------------
      340  total obs.
      162  obs. end on or before enter()
------------------------------------------------------------------------------
      178  obs. remaining, representing
      178  subjects
      112  failures in single failure-per-subject data
     2480  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        59

. sort id enum

. list id enum start stop event _t0 _t _d _st if id>6&id<10 & _st==1, sepby(id)
```

	id	enum	start	stop	event	_t0	_t	_d	_st
25.	7	1	0	18	0	0	18	0	1
29.	8	1	0	5	1	0	5	1	1
30.	8	2	5	18	0	5	18	0	1
33.	9	1	0	12	1	0	12	1	1
34.	9	2	12	16	1	12	16	1	1
35.	9	3	16	18	0	16	18	0	1

```
. stcox treat number size, strata(enum) vce(cluster id) efron
         failure _d:  event == 1
   analysis time _t:  stop
  enter on or after:  time start
                 id:  obs

Stratified Cox regr. -- Efron method for ties

No. of subjects      =          178             Number of obs   =       178
No. of failures      =          112
Time at risk         =         2480
                                                Wald chi2(3)    =      7.17
Log pseudolikelihood =   -315.99082             Prob > chi2     =    0.0665

                                 (Std. Err. adjusted for 85 clusters in id)

                           Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |     .71642    .147584    -1.62   0.105     .4784299    1.072796
      number |   1.127065   .0582599     2.31   0.021     1.018472    1.247238
        size |   .9915413   .0614766    -0.14   0.891     .8780828     1.11966
------------------------------------------------------------------------------
                                                          Stratified by enum
```

5. *Andersen and Gill (1982) also use counting process risk intervals, but they use unrestricted risk sets and assume that all events have a common baseline hazard function. Fit this model, again assuming that* `treat`, `number`, *and* `size` *have the same coefficients across events.*

```
. stcox c.treat c.number c.size, vce(cluster id) efron
         failure _d:  event == 1
   analysis time _t:  stop
  enter on or after:  time start
                 id:  obs

Cox regression -- Efron method for ties

No. of subjects      =          178             Number of obs   =       178
No. of failures      =          112
Time at risk         =         2480
                                                Wald chi2(3)    =     11.41
Log pseudolikelihood =   -449.98064             Prob > chi2     =    0.0097

                                 (Std. Err. adjusted for 85 clusters in id)

                           Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .6283318   .1678506    -1.74   0.082     .3722217     1.06066
      number |   1.191199   .0755395     2.76   0.006     1.051976    1.348848
        size |   .9572791   .0747412    -0.56   0.576      .821447    1.115572
------------------------------------------------------------------------------
```

(*Continued on next page*)

6. *In their model (3), Prentice, Williams, and Peterson (1981) use gap time with restricted risk sets and event-specific baseline hazards. Fit this model, assuming that* `treat`*,* `number`*, and* `size` *have the same coefficients across events.*

```
. stset stop, origin(start) failure(event=1) id(obs)
                id:  obs
     failure event:  event == 1
obs. time interval:  (stop[_n-1], stop]
 exit on or before:  failure
    t for analysis:  (time-origin)
            origin:  time start

------------------------------------------------------------------------------
      340  total obs.
      162  obs. end on or before enter()
------------------------------------------------------------------------------
      178  obs. remaining, representing
      178  subjects
      112  failures in single failure-per-subject data
     2480  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        59

. sort id enum

. list id enum start stop event _t0 _t _d _st if id>6&id<10 & _st==1, sepby(id)

     +----------------------------------------------------------+
     | id   enum   start   stop   event   _t0   _t   _d   _st |
     |----------------------------------------------------------|
 25. |  7      1       0     18       0     0   18    0     1 |
     |----------------------------------------------------------|
 29. |  8      1       0      5       1     0    5    1     1 |
 30. |  8      2       5     18       0     0   13    0     1 |
     |----------------------------------------------------------|
 33. |  9      1       0     12       1     0   12    1     1 |
 34. |  9      2      12     16       1     0    4    1     1 |
 35. |  9      3      16     18       0     0    2    0     1 |
     +----------------------------------------------------------+

. stcox treat number size, strata(enum) vce(cluster id) efron

         failure _d:  event == 1
   analysis time _t:  (stop-origin)
             origin:  time start
                 id:  obs

Stratified Cox regr. -- Efron method for ties

No. of subjects      =          178                Number of obs   =       178
No. of failures      =          112
Time at risk         =         2480
                                                   Wald chi2(3)    =     11.70
Log pseudolikelihood =   -358.96849                Prob > chi2     =    0.0085

                                   (Std. Err. adjusted for 85 clusters in id)
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .7565365   .1640954    -1.29   0.198     .4945398    1.157333
      number |    1.17122   .0600157     3.08   0.002     1.059305    1.294958
        size |   1.007443    .065196     0.11   0.909     .8874327    1.143682
------------------------------------------------------------------------------
                                                            Stratified by enum
```

7. *Compare and interpret the treatment effect estimates from steps 3 to 6.*

The estimated hazard ratios are 0.56 for total time semi-restricted, 0.72 for counting process, restricted, 0.63 for counting process unrestricted, and 0.76 for gap times, restricted. Only the total time semi-restricted estimate is nearly significant at the 5% level. The estimates can be interpreted as a 54% reduction in the hazard (largest effect size estimate) down to a 24% reduction in the hazard (smallest effect size estimate), controlling for number and maximum size of initial tumors.

16.2 Tower-of-London data

1. *Fit the two-level random-intercept model (random intercept for persons):*

$$\text{logit}\{\Pr(y_{ijk}\!=\!1 \mid \mathbf{x}_{ijk}, \zeta_{jk}^{(2)})\} \;=\; \beta_0 + \beta_1 x_{ijk} + \beta_2 g_{2ijk} + \beta_3 g_{3ijk} + \zeta_{jk}^{(2)}$$

where g_{2ijk} and g_{3ijk} are dummy variables for groups 2 and 3, respectively, and $\zeta_{jk}^{(2)} \sim N(0, \psi^{(2)})$ is independent of the covariates $\mathbf{x}_{ijk}$. Here and throughout the exercise, `level` *is treated as continuous.*

```
. use towerl, clear

. tabulate group, generate(g)

      GROUP |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        194       28.66       28.66
          2 |        294       43.43       72.08
          3 |        189       27.92      100.00
------------+-----------------------------------
      Total |        677      100.00

. rename g2 relatives

. rename g3 schizo

. gllamm dtlm level relatives schizo, i(id) link(logit) family(binomial) adapt

number of level 1 units = 677
number of level 2 units = 226

Condition Number = 4.4746865

gllamm model

log likelihood = -305.95923

------------------------------------------------------------------------------
        dtlm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       level |  -1.649218   .1934261    -8.53   0.000    -2.028326    -1.27011
   relatives |  -.1691618   .3343253    -0.51   0.613    -.8244274    .4861037
      schizo |  -1.023004    .393953    -2.60   0.009    -1.795137   -.2508701
       _cons |   -1.48306   .2836532    -5.23   0.000     -2.03901     -.92711
------------------------------------------------------------------------------


Variances and covariances of random effects
------------------------------------------------------------------------------


***level 2 (id)

    var(1): 1.6768915 (.66262435)
------------------------------------------------------------------------------

. estimates store mod0
```

The syntax for `xtmelogit` is

```
xtmelogit dtlm level relatives schizo || id:
```

The syntax for `xtlogit` is

```
xtset id
xtlogit dtlm level relatives schizo
```

2. *Fit the three-level random-intercept model (random intercepts for subjects and families):*

$$\text{logit}\{\Pr(y_{ijk}=1 \mid \mathbf{x}_{ijk}, \zeta_{jk}^{(2)}, \zeta_k^{(3)})\} \;=\; \beta_0 + \beta_1 x_{ijk} + \beta_2 g_{2ijk} + \beta_3 g_{3ijk} + \zeta_{jk}^{(2)} + \zeta_k^{(3)}$$

where $\zeta_{jk}^{(2)} \sim N(0, \psi^{(2)})$ *is independent of* $\zeta_k^{(3)} \sim N(0, \psi^{(3)})$ *and both random effects are assumed independent of* $\mathbf{x}_{ijk}$.

```
. gllamm dtlm level relatives schizo, i(id famnum) link(logit) family(binomial) adapt
number of level 1 units = 677
number of level 2 units = 226
number of level 3 units = 118

Condition Number = 4.2143936

gllamm model

log likelihood = -305.12037
```

dtlm	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
level	-1.648477	.1932181	-8.53	0.000	-2.027177	-1.269776
relatives	-.2487947	.3543655	-0.70	0.483	-.9433383	.4457489
schizo	-1.052438	.3999452	-2.63	0.009	-1.836317	-.26856
_cons	-1.48575	.2848124	-5.22	0.000	-2.043972	-.9275283

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id)

    var(1): 1.1370879 (.68588796)

***level 3 (famnum)

    var(1): .5690352 (.52168443)
------------------------------------------------------------------------------
. estimates store mod1
```

Subjects with schizophrenia perform significantly worse than unrelated healthy control subjects, whereas the healthy relatives of the subjects with schizophrenia do perform significantly worse than unrelated healthy control subjects (at the 5% level). Performance declines as the level of difficulty increases. There is more variability between subjects within families than between families after controlling for covariates.

The syntax for `xtmelogit` is

```
xtmelogit dtlm level relatives schizo || famnum: || id:
```

3. *Compare the models in steps 1 and 2 using a likelihood-ratio test, but retain the three-level model even if the null hypothesis is not rejected at the 5% level.*

```
. lrtest mod0 mod1
Likelihood-ratio test                                 LR chi2(1)  =      1.68
(Assumption: mod0 nested in mod1)                     Prob > chi2 =    0.1952
```

Since the random intercepts at the different levels are uncorrelated, we can divide the naïve p-value by 2 (see display 8.1, page 397) to obtain the correct asymptotic p-value of 0.10.

4. *Include a group (controls, relatives, schizophrenics) by level of difficulty interaction in the three-level model. Test the interaction using both a Wald test and a likelihood-ratio test.*

```
. generate lev_rel = level*relatives

. generate lev_sch = level*schizo

. gllamm dtlm level relatives schizo lev_rel lev_sch,
>   i(id famnum) link(logit) family(binomial) adapt
number of level 1 units = 677
number of level 2 units = 226
number of level 3 units = 118

Condition Number = 5.9640326

gllamm model

log likelihood = -301.8829
```

dtlm	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
level	-1.180727	.2643959	-4.47	0.000	-1.698933	-.6625202
relatives	-.4365425	.3705992	-1.18	0.239	-1.162904	.2898186
schizo	-1.611176	.5116238	-3.15	0.002	-2.61394	-.6084113
lev_rel	-.6126168	.3528075	-1.74	0.082	-1.304107	.0788733
lev_sch	-1.176511	.5209349	-2.26	0.024	-2.197525	-.1554972
_cons	-1.356816	.2797885	-4.85	0.000	-1.905192	-.8084408

```
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id)

    var(1): 1.2092826 (.69676348)

***level 3 (famnum)

    var(1): .53767723 (.48595677)
------------------------------------------------------------------------------
```

We obtain a Wald test by using `testparm`

```
. testparm lev_rel lev_sch
 ( 1)  [dtlm]lev_rel = 0
 ( 2)  [dtlm]lev_sch = 0
           chi2(  2) =    6.08
         Prob > chi2 =    0.0478
```

The interaction is significant at the 5% level according to the Wald test ($w = 6.09$, df $= 2$, $p = 0.048$). The corresponding likelihood-ratio test can be obtained using `lrtest`

```
. lrtest mod1 .
Likelihood-ratio test                                 LR chi2(2)  =      6.47
(Assumption: mod1 nested in .)                        Prob > chi2 =    0.0393
```

The likelihood-ratio statistic is 6.47 with two degrees of freedom, giving a p-value of 0.04.

For schizophrenics, performance declines faster with increasing level of difficulty than for controls ($z = -2.26$, $p = 0.024$).

5. *For the model in step 4, obtain predicted marginal or population-averaged probabilities using* `gllapred`*. (This requires fitting the model in* `gllamm`*.) Plot the probabilities against the levels of difficulty with different curves for the three groups.*

```
. gllapred prob, mu marg
(mu will be stored in prob)
. twoway (line prob level if group==1, sort)
>   (line prob level if group==2, sort lpatt(longdash))
>   (line prob level if group==3, sort lpatt(shortdash)),
>   xtitle(Level of difficulty) ytitle(Probability)
>   legend(order(1 "Controls" 2 "Relatives" 3 "Schizophrenics") row(1))
>   xlabel(-1 "Low" 0 "Medium" 1 "High")
```

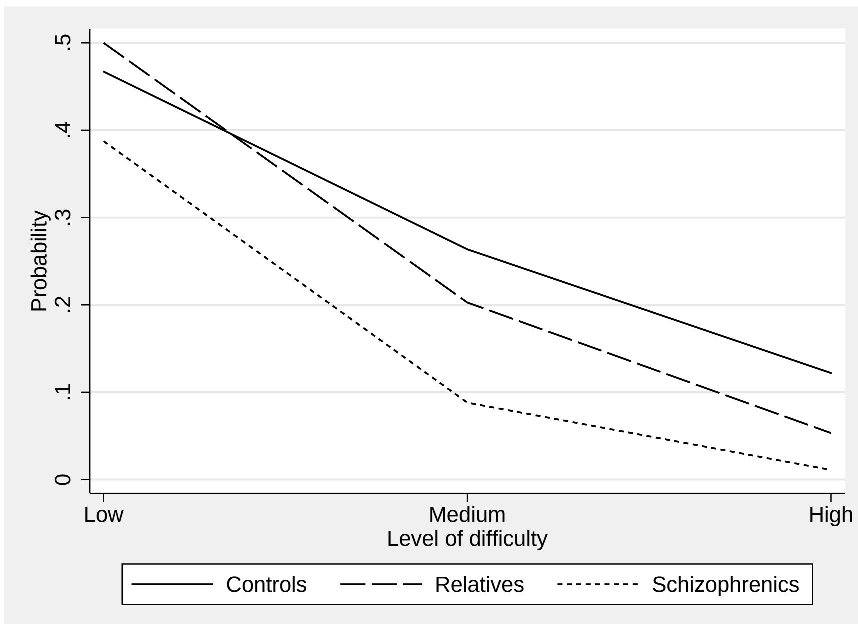

Figure 6: Predicted marginal probabilities as a function of level of difficulty for the three groups.