

# Recent Developments in Multilevel Modeling

Roberto G. Gutierrez

Director of Statistics  
StataCorp LP

2007 North American Stata Users Group Meeting, Boston



1. What's new in Stata 10
2. One-level models
3. Alternate covariance structures
4. A two-level model
5. The Laplacian approximation
6. A crossed-effects model
7. Concluding remarks

- New commands `xtmelogit` and `xtmepoisson`
- Mixed effects for binary and count responses
- They work just like `xtmixed` does
- Random intercepts and random coefficients
- You can have multiple levels of nested random effects
- Various predictions, including random effects and their standard errors
- We'll be discussing binary responses and `xtmelogit`

For a series of  $i = 1, \dots, M$  independent panels, let

$$P(y_{ij} = 1 | \mathbf{u}_i) = H(\mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{u}_i)$$

where

there are  $j = 1, \dots, n_{ij}$  observations in panel  $i$

$\mathbf{x}_{ij}$  are the  $p$  covariates for the fixed effects

$\boldsymbol{\beta}$  are the fixed effects

$\mathbf{z}_{ij}$  are the  $q$  covariates for the random effects

$\mathbf{u}_i$  are the random effects, specific to panel  $i$

$\mathbf{u}_i$  are normal with mean  $\mathbf{0}$  and variance matrix  $\boldsymbol{\Sigma}$

$H()$  is the logistic cdf

- You can also think of this model in terms of a latent response  $y_{ij} = I(y_{ij}^* > 0)$  where

$$y_{ij}^* = \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{u}_i + \epsilon_{ij}$$

- The errors  $\epsilon_{ij}$  are logistic-distributed with mean zero and variance  $\pi^2/3$ , independent of  $\mathbf{u}_i$

- Random effects are not directly estimated, but instead characterized by the elements of  $\Sigma$ , known as *variance components*
- You can, however, “predict” random effects
- As such, you fit this model by estimating  $\beta$  and the variance components in  $\Sigma$
- A maximum-likelihood solution requires integrating out the distribution of  $\mathbf{u}_i$ .
- A tricky proposition in nonlinear models such as logit

## Example

- 1989 Bangladesh fertility survey (Huq and Cleland 1990)
- Ng et al. (2006) analyze data on 1,934 women, who were polled on their use of contraception
- Data were collected from 60 districts containing urban and rural areas
- Covariates include age, urban/rural area, and indicators for number of children
- Among other things, we wish to assess a district effect on contraception use

- For woman  $j$  in district  $i$ , consider this model for

$$\pi_{ij} = P(c\_use_{ij} = 1)$$

$$\begin{aligned} \text{logit}(\pi_{ij}) = & \beta_0 + \beta_1 \text{urban}_{ij} + \beta_2 \text{age}_{ij} + \\ & \beta_3 \text{child1}_{ij} + \beta_4 \text{child2}_{ij} + \beta_5 \text{child3}_{ij} + u_i \end{aligned}$$

- The  $u_i$  represent 60 district-specific random effects
- You can use `xtlogit` (option `re`) to fit this model and estimate  $\sigma_u^2$ , the variance of the  $u_i$
- `xtlogit` will also give an LR test for  $H_o: \sigma_u^2 = 0$ , by comparing log likelihoods with `logit`
- You could also use `xtmelogit` on this model

- Introducing a random coefficient, we now consider

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 \text{urban}_{ij} + \mathcal{F}_{ij} + u_i + v_i \text{urban}_{ij}$$

- $\mathcal{F}_{ij}$  is shorthand for the fixed-effects specification on age and children
- This model allows for distinct random effects for urban and rural areas within each district
- For rural areas in district  $i$ , the effect is  $u_i$
- For urban areas,  $u_i + v_i$
- You need `xtmelogit` to fit this model

```
. xtlogit c_use urban age child* || district: urban
```

```
Refining starting values:
```

```
(output omitted)
```

```
Performing gradient-based optimization:
```

```
(output omitted)
```

```
Mixed-effects logistic regression
```

```
Group variable: district
```

```
Number of obs      =      1934
```

```
Number of groups   =         60
```

```
Obs per group: min =          2
```

```
                  avg =      32.2
```

```
                  max =      118
```

```
Integration points =    7
```

```
Log likelihood = -1205.0025
```

```
Wald chi2(5)      =      97.30
```

```
Prob > chi2       =      0.0000
```

c_use	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
urban	.7143927	.1513595	4.72	0.000	.4177336 1.011052
age	-.0262261	.0079656	-3.29	0.001	-.0418384 -.0106138
child1	1.128973	.1599346	7.06	0.000	.8155069 1.442439
child2	1.363165	.1761804	7.74	0.000	1.017857 1.708472
child3	1.352238	.1815608	7.45	0.000	.9963853 1.70809
_cons	-1.698137	.1505019	-11.28	0.000	-1.993115 -1.403159

```
--more--
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
district: Independent				
sd(urban)	.5235464	.203566	.2443374	1.121813
sd(_cons)	.4889585	.087638	.3441182	.6947624

LR test vs. logistic regression:       $\chi^2(2) = 47.05$       Prob >  $\chi^2 = 0.0000$

Note: LR test is conservative and provided only for reference.

- As with `logit`, option `or` will give odds ratios
- Use option `variance` for variances instead of standard deviations of random effects
- LR test comparing to standard `logit` is at the bottom, along with a note telling you the  $p$ -value is conservative

- Evaluating the log likelihood requires integrating out the random effects
- The default method used by `xtmelogit` is adaptive Gaussian quadrature (AGQ) with seven quadrature points per level
- AGQ is computationally intensive
- Previous methods, such as PQL and MQL, avoided the integration altogether (Breslow and Clayton 1993)
- PQL and MQL can be severely biased (Rodriguez and Goldman 1995)
- Also, being quasi-likelihood, their use prohibits LR tests

- Implicit in our previous model was the default independent covariance structure

$$\Sigma = \text{Var} \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}$$

- Assuming  $\text{Cov}(u_i, v_i) = 0$  means you are also assuming  $\text{Var}(u_i + v_i) > \text{Var}(u_i)$
- Are urban areas really more variable than rural areas?
- Even worse, what if we change the coding of the random effects? Codings are not arbitrary here
- Option `covariance(unstructured)` will include this covariance in the model

```
. xtlogit c_use urban age child* || district: urban, cov(un) var
      (output omitted)
```

```
Mixed-effects logistic regression
Group variable: district
```

```
Number of obs      =      1934
Number of groups   =         60
Obs per group: min =          2
                  avg =      32.2
                  max =      118
```

```
Integration points =    7
Log likelihood = -1199.315
```

```
Wald chi2(5)      =      97.50
Prob > chi2       =      0.0000
```

c_use	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
urban	.8157872	.1715519	4.76	0.000	.4795516	1.152023
age	-.026415	.008023	-3.29	0.001	-.0421398	-.0106902
child1	1.13252	.1603285	7.06	0.000	.818282	1.446758
child2	1.357739	.1770522	7.67	0.000	1.010724	1.704755
child3	1.353827	.1828801	7.40	0.000	.9953882	1.712265
_cons	-1.71165	.1605617	-10.66	0.000	-2.026345	-1.396954

```
--more--
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
district: Unstructured				
var(urban)	.6663222	.3224715	.2580709	1.7204
var(_cons)	.3897435	.1292459	.2034723	.7465388
cov(urban,_cons)	-.4058846	.1755418	-.7499403	-.0618289

LR test vs. logistic regression:      chi2(3) =    58.42    Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

```
. estimates store corr
```

```
. lrtest no_corr corr
```

```
Likelihood-ratio test
```

```
(Assumption: no_corr nested in corr)
```

```
LR chi2(1) =       11.38
```

```
Prob > chi2 =       0.0007
```

- We can now estimate the variance of the random effects for urban areas as

$$\text{Var}(u_i + v_i) = \sigma_u^2 + \sigma_v^2 + 2\sigma_{uv}$$

- If you did this, you would get  $\text{Var}(u_i + v_i) = 0.244$ , which is actually less than  $\text{Var}(u_i) = 0.390$
- Better still, if you want to directly compare rural areas to urban areas, recode your random effects
- The unstructured covariance structure will ensure an equivalent model under alternate codings of random-effects variables
- Also, predictions of random effects will be what you want

```

. gen byte rural = 1 - urban
. xtmelogit c_use urban rural age child*, nocons || district: urban rural,
> nocons cov(un) var
  (output omitted)
Mixed-effects logistic regression
Group variable: district

Number of obs      =      1934
Number of groups   =         60
Obs per group: min =         2
                  avg =      32.2
                  max =      118

Integration points =    7
Log likelihood = -1199.315
  (output omitted)

Wald chi2(6)      =      120.24
Prob > chi2       =      0.0000

```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
district: Unstructured				
var(urban)	.2442916	.1450648	.0762869	.7822893
var(rural)	.3897431	.1292457	.2034722	.7465379
cov(urban,rural)	-.0161406	.105746	-.2233989	.1911177

```
LR test vs. logistic regression:    chi2(3) =    58.42    Prob > chi2 = 0.0000
```

```
Note: LR test is conservative and provided only for reference.
```

- You've seen Independent and Unstructured in action
- Also available are Identity and Exchangeable
- You can combine these to form blocked-diagonal structures
- Such structures can reduce the number of estimable parameters
- For example, consider a random effects specification of the form

```
... || district: child1 child2, nocons cov(ex) || district: child3, nocons
```

as an alternative to a  $3 \times 3$  unstructured variance matrix

## Example

- The Tower of London (Rabe-Hesketh et al. 2001)
- Study of cognitive abilities of patients with schizophrenia
- Cognitive ability was measure as successful completion of the Tower of London, a computerized task (binary variable `dt1m`)
- 226 subjects, all but one tested at three difficulty levels
- Subjects were not only patients (`group==3`), but relatives (`group==2`) and nonrelated controls (`group==1`)
- We can thus propose a model having random effects shared among relatives (variable `family`) and subject-specific effects nested within families

```
. xi: xtmelogit dtlm difficulty i.group || family: || subject:, or variance
i.group      _Igroup_1-3      (naturally coded; _Igroup_1 omitted)
(output omitted)
Mixed-effects logistic regression      Number of obs      =      677
```

Group Variable	No. of Groups	Observations per Group			Integration Points
		Minimum	Average	Maximum	
family	118	2	5.7	27	7
subject	226	2	3.0	3	7

```
Log likelihood = -305.12043      Wald chi2(3)      =      74.89
                                Prob > chi2      =      0.0000
```

dtlm	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
difficulty	.192337	.0371622	-8.53	0.000	.131704	.2808839
_Igroup_2	.7798295	.2763766	-0.70	0.483	.3893394	1.561964
_Igroup_3	.3491338	.1396499	-2.63	0.009	.1594117	.7646517

```
--more--
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
family: Identity var(_cons)	.569182	.5216584	.0944322	3.430694
subject: Identity var(_cons)	1.137931	.6857497	.3492672	3.707441

LR test vs. logistic regression:       $\chi^2(2) = 17.54$     Prob >  $\chi^2 = 0.0002$

Note: LR test is conservative and provided only for reference.

- `xtmelogit`, by default, uses AGQ which can be intensive with large datasets or high-dimensional models
- Computation time is roughly on the order of

$$T \sim p^2 \{M + M(N_Q)^{q_t}\}$$

where

$p$  is the number of estimable parameters

$M$  is the number of lowest-level (smallest) panels

$N_Q$  is the number of quadrature points

$q_t$  is the total dimension of the random effects (all levels)

- The real killer is  $(N_Q)^{q_t}$

- Ideally, you want enough quadrature points such that adding more points doesn't change much
- In complex models, this can very time consuming, especially during the exploratory phase of the analysis
- Sometimes you just want quicker results, and you may be willing to give up a bit of accuracy
- Use option `laplace`, equivalent to  $N_Q = 1$
- The computational benefit is clear – one raised to any power equals one

```
. xi: xtlogit dtlm level i.group || family: || subject:, or variance laplace
i.group      _Igroup_1-3      (naturally coded; _Igroup_1 omitted)
(output omitted)
Mixed-effects logistic regression      Number of obs      =      677
```

Group Variable	No. of Groups	Observations per Group			Integration Points
		Minimum	Average	Maximum	
family	118	2	5.7	27	1
subject	226	2	3.0	3	1

```
Log likelihood = -306.51035      Wald chi2(3)      =      76.09
                                Prob > chi2      =      0.0000
```

dtlm	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
level	.2044132	.0377578	-8.60	0.000	.1423248	.2935872
_Igroup_2	.7860452	.2625197	-0.72	0.471	.4084766	1.512613
_Igroup_3	.3575718	.1354592	-2.71	0.007	.1701774	.7513194

```
--more--
```

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
family: Identity				
var(_cons)	.522942	.4704255	.0896879	3.04911
subject: Identity				
var(_cons)	.7909329	.5699273	.1926568	3.247095

LR test vs. logistic regression:       $\chi^2(2) = 14.76$       Prob >  $\chi^2 = 0.0006$

Note: LR test is conservative and provided only for reference.

Note: log-likelihood calculations are based on the Laplacian approximation.

- Odds ratios and their standard errors are well approximated by Laplace
- Variance components exhibit bias, particularly at the lower (subject) level
- Model log-likelihoods and comparison LR test are in fair agreement
- These behaviors are fairly typical
- If anything, it shows that you can at least use `laplace` while building your model

One further advantage of `laplace` is that it permits you to fit crossed-effects models, which will have high-dimension

### Example

- School data from Fife, Scotland (Rabe-Hesketh and Skrondal 2005)
- Attainment scores at age 16 for 3,435 students who attended any of 148 primary schools and 19 secondary schools
- We are interested in whether the attainment score is greater than 6
- We want random effects due to primary school and secondary school, but these effects are not nested

- Consider the model

$$\text{logit}\{\text{Pr}(\text{attain}_{ijk} > 6)\} = \beta_0 + \beta_1 \text{sex}_{ijk} + u_i + v_j$$

for student  $k$  who attended primary school  $i$  and secondary school  $j$

- Since there is no nesting, you can use the level designation `_all:` to treat the entire data as one big panel
- Use factor notation `R. varname` to mimic the creation of indicator variables identifying schools
- However, notice that we can treat one set of effects as nested within the entire data

```
. xtlogit attain_gt_6 sex || _all:R.sid || pid:, or variance
```

Note: factor variables specified; option laplace assumed

(output omitted)

```
Mixed-effects logistic regression                Number of obs      =      3435
```

Group Variable	No. of Groups	Observations per Group			Integration Points
		Minimum	Average	Maximum	
_all	1	3435	3435.0	3435	1
pid	148	1	23.2	72	1

```
Log likelihood = -2220.0035                Wald chi2(1)      =      14.28
                                           Prob > chi2       =      0.0002
```

attain_gt_6	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
sex	1.32512	.0986968	3.78	0.000	1.145135	1.533395

```
--more--
```

Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]	
<b>_all: Identity</b>					
	var(R.sid)	.1239739	.0694743	.0413354	.3718252
<b>pid: Identity</b>					
	var(_cons)	.4520502	.0953867	.298934	.6835937

LR test vs. logistic regression:       $\chi^2(2) = 195.80$       Prob >  $\chi^2 = 0.0000$

Note: LR test is conservative and provided only for reference.

Note: log-likelihood calculations are based on the Laplacian approximation.

- `xtmelogit` and `xmepoisson` are new to Stata 10
- We discussed `xtmelogit` – the same holds true for `xmepoisson`
- Computations can get intensive
- The Laplacian approximation is a quicker alternative
- You can fit crossed-effects models, and large ones with creative nesting
- Work in this area is ongoing

- Breslow, N. E. and D. G. Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88: 9–25.
- Huq, N. M. and J. Cleland. 1990. *Bangladesh Fertility Survey 1989 (Main Report)*. National Institute of Population Research and Training.
- Ng, E. S. W., J. R. Carpenter, H. Goldstein, and J. Rasbash. 2006. Estimation in generalised linear mixed models with binary outcomes by simulated maximum likelihood. *Statistical Modelling* 6: 23–42.
- Rabe-Hesketh, S., S. R. Touloupoulou, and R. M. Murray. 2001. Multilevel modeling of cognitive function in schizophrenics and their first degree relatives. *Multivariate Behavioral Research* 36: 279–298.
- Rabe-Hesketh, S. and A. Skrondal. 2005. *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Press.
- Rodriguez, G. and N. Goldman. 1995. An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* 158: 73–89.