An Introduction to Multi-Level Models (Using Stata)

European University Institute, May 23–27, 2011

Professor Kenneth Benoit
Methodology Institute
London School of Economics
http://www.kenbenoit.net/

Preliminary Version: May 19, 2011

Objectives and Learning Outcomes

Multilevel models are a class of models that are used for data that are clustered into hierarchically organized groups, violating the conditional independence assumption about the error term from the classical linear regression model. For example, political data is frequently observed for individuals that are grouped by region or country, and there may be slightly different effects for individuals according to which region or country they are from. Multilevel models not only provide a means to correctly estimate causal models when data is hierarchically clustered, but also provide more direct ways to investigate the effects of different levels themselves on the causal process.

Prerequisites

Students in this course should already understand the linear regression model and basic concepts of statistical inference. Day 2 will provide a review of these topics, however. Some familiarity with Stata is required, since this software package will be used for examples and for the class exercises. Students are welcome to implement the same models in R (see below).

Logistics

Meetings. Classes will meet five days for one two-hour sessions each day, from 11-13:00, from Monday May 23, 2011 through Friday May 27, 2011. In addition, on Days 2 through 5 a lab session will be held from 09:30-10:30 before the course.

Computer Software. Stata 11 will be used for this course. It would be an excellent idea to begin familiarizing yourselves with the Stata `XT` manual (*Longitudinal/Panel Data*). Students are also welcome to use the R statistical package for homeworks. Several libraries are available for this, including `multilevel`, `nmle`, and `lme4`.

Grading. The course is not graded, nor will I issue homework problem sets to be completed. Instead, course work will be done interactively, in class, on days 2–5. Your writeup of these class sessions should be emailed to me as record of your participation.

Datasets, code, and slides will be available from my web page at http://www.kenbenoit.net/mlm/.

Texts

A choice has been very deliberately made in this course to keep the reading material light in both quantity and difficulty level. We also chose to keep the methodological discussion of the models very close to the Stata implementation of these models. Consequently, most readings will be from a single text:

• S. Rabe-Hesketh and A. Skrondal. *Multilevel and Longitudinal Modeling Using Stata*. Stata Press, 2nd edition, 2008.

It is also highly recommended that you spend some time Reading The Fine Manual (*RTFM!*) for the Stata commands we will use in this course. These are found in the Stata 11 `XT` manual, which can be found in its electronic entirety through the Stata menu `Help → PDF Documentation`. Your reading should focus on the following commands:

- `xtmixed` Multilevel, mixed effects linear regression; see also the entry for `xtmixed postestimation`
- `xtreg` Fixed-, between-, and random-effects linear models; see also the entry for `xtreg postestimation`
- `xtmelogit` Multilevel, mixed effects logistic regression
- `xtmepoisson` Multilevel, mixed effects Poisson regression

Although it will not be required in this course, you may also wish to download and read about the `gllamm` library (Generalized Linear Latent And Mixed Models) available from http://www.gllamm.org. GLAMM was used with earlier versions of Stata for generalized mixed effects models, but most of these have now been incorporated into Stata 11.

Additional recommended readings are listed at the end of this handout and linked to each day of the course. One excellent additional text that I would suggest you purchase and read as a general reference as well as a text for multi-level models is:

- Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Schedule

1. **Day 1: Introduction and Motivation for multi-level models.**

   Discussion of the problem of multi-level units, introduction to the Rabe-Hesketh and Skrondal notation using *i* and *j* to subscript levels. Discussion of data organization for multilevel data, especially the “long” and “wide” formats, using examples. Simple models will be demonstrated using OLS to show that data problems with multilevel structures can nonetheless be estimated using standard methods, and we will interepret and discuss these results to suggest why MLMs might be needed. I will also introduce the variance-components model and preview fixed- versus random-effects models.

   Required Reading:

   Rabe-Hesketh & Skrondal (2008, Chs. 1–2); Stata manual for reshape.

   Recommended Reading:

Franzese (2005); Gelman (2006); Austin, Goel & van Walraven (2001).

Homework:

TBA

2. **Day 2: Estimating models with multi-level data.**

   Day 2 will start by revisiting the assumptions of the classical linear regression model, focusing on the assumptions that apply to the error term. A full coverage of the notation from R&S will be covered, and the full variance-components model will be used to decompose the variation in regression models where different hierarchical levels provide separate sources of variation. Intra-class correlation will be introduced and explained.

   Required Reading:

   Continue with Rabe-Hesketh & Skrondal (2008, Chs. 1–2); Stata `XT` manual, relevant commands.

   Recommended Reading:

   Steenbergen & Jones (2002); Austin, Goel & van Walraven (2001); Snijders & Bosker (1999); Goldstein (2003).

   Homework:

   TBA

3. **Day 3: Random-intercept models.**

   The focus for Day 3 will be to continue on the presentation of the general multilevel model structure from Day 2 but to focus specifically on the random-intercept model with covariates as described in Chapter 3 of Rabe-Hesketh & Skrondal (2008). The discussion centres on the disctinction between within-cluster and between-cluster covariate effects, and the problem of omitted cluster-level covariates and endogeneity.

   Required Reading:

   Rabe-Hesketh & Skrondal (2008, Ch. 3).

   Recommended Reading:

   Austin, Goel & van Walraven (2001); Snijders & Bosker (1999).

   Homework:

   TBA

4. **Day 4: Random-coefficient models.**

   Day 4 adds the possibility of random slopes to the previous models of random intercepts, so that the effects of covariates may differ across clusters.

   Required Reading:

   Rabe-Hesketh & Skrondal (2008, Ch. 4)

   Recommended Reading:

   Austin, Goel & van Walraven (2001); Snijders & Bosker (1999).

   Homework:

   TBA

5. **Day 5: Extensions of the multi-level model.**

This session will cover mixed models, `xtmelogit`, `xtmepoisson` for binary and count data. It will also introduce, but not present comprehensively, multilevel models for longitudinal and panel-structured data, as well as cross-classified data.

Required Reading:

Rabe-Hesketh & Skrondal (2008, Chs. 6, 9).

Recommended Reading:

McMahon & Heath (1992).

Homework:

TBA

Bibliography

Austin, Peter C., Vivek Goel & Carl van Walraven. 2001. "An introduction to multilevel regression models." *Canadian Journal of Public Health* 92(2, March-April):150–154.

Franzese, Robert. 2005. "Empirical Strategies for Various Manifestations of Multilevel Data." *Political Analysis* 13(4):430–436.

Gelman, Andrew. 2006. "Multilevel modeling: what it can and can't do." *Technometrics* 48:241–251.

Gelman, Andrew & Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical*

Models. Analytical Methods for Social Research Cambridge: Cambridge University Press.

Goldstein, Harvey. 2003. *Multilevel Statistical Models*. Oxford: Oxford University Press.

McMahon, Dorren & Anthony Heath. 1992. "Class and Party in Britain: Preliminary Results with a Multilevel Logit Model." *Multilevel Modelling Newsletter* 4(3):5-8(3, November):5–8.

Rabe-Hesketh, Sophia & Anders Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata*. 2nd ed. Stata Press.

Snijders, T.A.B. & R.J. Bosker. 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Stata Corp. 2007. *[XT] Longitudinal/Panel Data: Release 10*. Stata Press. Steenbergen, Marco R. & Bradford S. Jones. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46(1):218–237.