POLITICAL SCIENCE 676:02: MULTILEVEL MODELING

Department of Political Science ● Stony Brook University

Spring 2007 ● Thursdays, 10:00-1:00 ● SBS N-702

Instructor:	Brandon Bartels
Email:	brandon.bartels@stonybrook.edu
Office:	SBS S-713
Office phone:	632-7452
Office hours:	Mondays and Wednesdays, 1:30-2:30, or by appointment
Webpage:	http://ms.cc.sunysb.edu/~bbartels *Note*: I will post course materials, data, and announcements on Blackboard.

In this course, you will master the foundations of a methodology consisting of a class of models falling under the "multilevel modeling" umbrella. Multilevel data structures consist of more than one level of analysis, with each level consisting of different units of analysis. In the most common multilevel structure, data are hierarchical and consist of a nesting structure. In a two-level hierarchical structure, units of analysis from the lowest level (level one) are nested within units from level two (for example, voters [level-one units] nested within congressional districts [level-two units]). In a three-level model, a third level of analysis is present, within which level-two units are nested (e.g., voters nested within congressional districts nested within states [level-three units]). We will examine additional structures such as cross-classified data, where the data are non-nested and non-hierarchical. In addition, many types of well-known data structures are special forms of multilevel structures, including panel data, time series cross sectional (TSCS) data, and item response models.

Importantly, all varieties of multilevel models (including those discussed above) share common types of statistical and methodological issues, and moreover, they allow unique opportunities for modeling various types of political and behavioral processes. We will examine a wide range of multilevel modeling topics in this course. The broad course sequence is as follows:

I. Methodological Motivation: Multilevel Processes and Structures, Latent Constructs, Unobserved Heterogeneity, and Pooling
II. Model Specification: Hierarchical Generalized Linear Modeling and the Random Coefficient Specification
III. Estimation (Maximum Likelihood and Bayesian estimation), Model Checking, and Diagnostics
IV. Applications

REQUIRED BOOKS

The following two books are required for the course. I will not place orders of the books at the University bookstore, as you will be much better off purchasing these online.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC. [Note: Hereinafter, "S&R-H"]

Rabe-Hesketh, Sophia, and Anders Skrondal. 2005. *Multilevel and Longitudinal Modeling Using Stata*. Stata Press. [Available at Stata Press (www.statapress.com)] [Note: Hereinafter, "R-H&S"]

ADDITIONAL READINGS

In addition to the required readings, we'll draw upon some of the following readings, or portions thereof, throughout the course:

Bafumi, Joseph, and Andrew Gelman. 2006. "Fitting Multilevel Models When Predictors and Group Effects Correlate." Presented at the Annual Meeting of the MPSA.

Bafumi, Joseph, Luke Keele, David K. Park, and Boris Shor. 2006. "Examining Time Series Cross Sectional Data with Bayesian Multilevel Models." Working Paper.

Beck, Nathaniel. 2001. "Time-Series—Cross-Sectional Data: What Have We Learned in the Past Few Years?" *Annual Review of Political Science* 4:271-93.

Bowler, Shaun, Todd Donovan, and Robert Hanneman. 2003. "Art for Democracy's Sake? Group Membership and Political Engagement in Europe." *JOP* 65:1111-29.

Cho, Wendy K. Tam, James G. Gimpel, and Joshua J. Dyck. 2006. "Residential Concentration, Political Socialization, and Voter Turnout." *JOP* 68:156-67.

Congdon, Peter. 2003. *Applied Bayesian Modelling*. Wiley.

Congdon, Peter. 2005. *Bayesian Models for Categorical Data*. Wiley.

Gelman, Andrew, John Carlin, Hal Stern, and Donald Rubin. 2003. *Bayesian Data Analysis*. Chapman & Hall/CRC.

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

Gelman, Andrew, and Iain Pardoe. 2006. "Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models." *Technometrics* 48:241-51.

Gelman, Andrew, and Iain Pardoe. 2007. "Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components." *Sociological Methodology* (Forthcoming).

Heckman, James J. 1981. "Heterogeneity and State Dependence." In *Studies in Labor Markets*, ed. S. Rosen. Chicago, IL: University of Chicago Press.

Hox, Joop. 2002. *Multilevel Analysis*. Lawrence Erlbaum Associates.

Hsiao, Cheng. 2003. *Analysis of Panel Data* (Second Edition). Cambridge.

Martin, Andrew D. 2001. "Congressional Decision Making and the Separation of Powers." *APSR* 95:361-78.

Martin, Andrew D. 2003. "Bayesian Inference for Heterogeneous Event Counts." *Sociological Methods & Research* 32:30-63.

Park, Jong Hee, and Nathan Jensen. 2007. "Electoral Competition and Agricultural Support in OECD Countries." *AJPS* (Forthcoming, April).

Rabe-Hesketh, Sophia, Anders Skrondal, and Andrew Pickles. 2002. "Reliable Estimation of Generalized Linear Mixed Models Using Adaptive Quadrature." *Stata Journal* 2:1-21.

Rahn, Wendy M., and Thomas J. Rudolph. 2005. "A Tale of Political Trust in American Cities." *POQ* 69:530-60.

Raudenbush, Stephen, and Anthony Bryk. 2002. *Hierarchical Linear Modeling*. Sage.

Rodriguez, German, and Noreen Goldman. 2001. "Improved Estimation Procedures for Multilevel Models with Binary Response: A Case-Study." *Journal of the Royal Statistical Society, Series A* 164:339-55.

Rudolph, Thomas J. 2003. "Institutional Context and the Assignment of Political Responsibility." *JOP* 65:190-215.

Sastry, Narayan. 1997. "A Nested Frailty Model for Survival Data, With an Application to the Study of Child Survival in Northeast Brazil." *JASA* 92:426-35.

Steenbergen, Marco, and Bradford Jones. 2002. "Modeling Multilevel Data Structures." *AJPS* 46:218-37.

Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach." *AJPS* 42:1233-59.

Western, Bruce, and Simon Jackman. 1994. "Bayesian Inference for Comparative Research." *APSR* 88:412-23.

Wilson, Sven, and Daniel M. Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis* (Forthcoming).

Wooldridge, Jeffrey M. 2005. "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics* 20:39-54.

Zorn, Christopher. 2000. "Modeling Duration Dependence." *Political Analysis* 8:367-80.

SOFTWARE

We'll rely on two software programs: *Stata 9* and *WinBUGS*. We will use Stata's built-in commands for estimating multilevel models (the "xt" family and the "xtmixed" command) as well as *GLLAMM*, which is an add-on package to Stata (created by Sophia Rabe-Hesketh, the co-author of both required textbooks) that estimates multilevel models. All of the worked examples in the Rabe-Hesketh and Skrondal book (the second one in the required readings list) use Stata. *WinBUGS*, which can be downloaded for free, estimates models via the Bayesian / MCMC approach (Gibbs sampling). While we'll primarily rely on these two programs, I will show you differences and similarities between these programs and others such as *R* and *HLM*.

REQUIREMENTS (500 points total)
Class participation: Students are expected to complete all of the assigned readings, to attend each class, and to participate in class sessions (50 points).

Problem sets: Four major problem sets (100 points each) will constitute the supermajority of your responsibilities. Each problem set will require you to demonstrate your understanding of the material and the ability to make appropriate interpretations. The due dates, also noted in the Course Outline section of the syllabus, are as follows:

Problem Set 1: Thursday, February 22
Problem Set 2: Thursday, March 22
Problem Set 3: Thursday, April 19
Problem Set 4: Thursday, May 10

Two important notes on problem sets:

- All problem sets will be due at *the beginning of class* on the day they are due. *I will not accept late problem sets*. Also, these are in-depth, intense assignments, so you will not want to wait until the last minute to start them. I strongly suggest working on the assignments progressively as we get through the relevant material.
- I encourage you to work with your classmates on the problem sets. Collaboration can be beneficial for mastering the material. However, ***you must do your own work***. That is, while you can work together, ***the final product that you hand in must be your own work***.

Class presentation (50 points): Each student will give a presentation to the class on an advanced multilevel modeling topic, preferably one in which you have a particular interest. A list of advanced topics is provided at the end of the syllabus. You are more than welcome to choose a topic that is not on this list as long as you get my prior approval. I would like to spread out the presentations throughout the semester, so please let me know your topic ASAP so that I can plan accordingly. For the actual presentation, students are expected to do some background reading on the topic, including some outside readings. The presentation, and associated discussion, should be around 30 minutes, so you should plan a presentation that is about 15-20 minutes.

ACADEMIC HONESTY
All of the work you do in this class is expected to be your own. Cheating and plagiarism are the grandest offenses a student can commit. Therefore, I will handle cases of cheating and plagiarism according to university policy. The bottom line: *Don't cheat*! It's simply not worth risking your academic career for!

STUDENTS WITH DISABILITIES
Any student who feels that they may need assistance should inform the instructor at the beginning of the semester so that adequate arrangements can be made.

COURSE OUTLINE AND SCHEDULE

I. *Methodological Motivation: Multilevel Processes and Structures, Latent Constructs, Unobserved Heterogeneity, and Pooling* - S&R-H, Ch. 1; Steenbergen and Jones; portions of Hox; Raudenbush and Bryk; and Hsiao **[1/25 & 2/1]**

II. *Model Specification: Hierarchical Generalized Linear Modeling and the Random Coefficient Specification* – S&R-H, Ch. 2-4 **[2/1, 2/8, 2/15]**

***** Problem Set 1 due Thu, 2/22***

III. *Estimation (Maximum Likelihood and Bayesian estimation), Model Checking, and Diagnostics* – S&R-H, Ch. 5-8; portions of both Congdon books; portions of Gelman et al. and Gelman and Hill; Gelman and Pardoe 2006, 2007; Bafumi and Gelman; Rodriguez and Goldman; Western; Western and Jackman; Rabe-Hesketh et al. **[2/15, 2/22, 2/29, 3/8]**

IV. *Applications: Specification, Estimation, Evaluation, Interpretation, and Post-estimation*

A. Models for continuous responses – R-H&S, Ch. 1-3; Bowler et al.; Rahn and Rudolph **[3/15]**

***** Problem Set 2 due Thu, 3/22***

B. Models for dichotomous responses – R-H&S, Ch. 4; S&R-H, Ch. 9; Martin 2001; Cho et al. **[3/22]**
C. Models for ordinal and count data - R-H&S, Ch. 5-6; S&R-H, Ch. 10-11; Martin 2003 **[3/29]**
D. Models for nominal responses - S&R-H, Ch. 13; Rudolph **[4/12]**

***** Problem Set 3 due Thu, 4/19***

E. Models for duration data - S&R-H, Ch. 12; Sastry; Zorn **[4/19]**
F. Models for panel and TSCS data – Hsiao, TBA; Beck; Bafumi et al.; Western; Bafumi and Gelman; Wilson and Butler; Heckman; Wooldridge **[4/26]**
G. Cross-classified, non-nested models – R-H&S, Ch. 8; S&R-H, pp. 60-63; Park and Jensen **[5/3]**

***** Problem Set 4 due Thu, 5/10***

TOPICS FOR STUDENT PRESENTATIONS

- GEE versus random intercept specifications (see S&R-H)
- Average partial effects versus conditional effects (e.g., Gelman and Pardoe 2007; S&R-H)
- Post-estimation methods for multilevel models using MCMC and ML (e.g., Gelman and Pardoe 2007; S&R-H; R-H&S)
- Goodness of fit in multilevel models (e.g., Gelman and Pardoe 2006; S&R-H)
- Pooling in multilevel models: Conceptual issues and measures (e.g., Gelman and Pardoe 2006)

- Comparing estimation algorithms for multilevel models with nonlinear responses: Quadrature v. PQL/MQL v. MCMC (e.g., Rodriguez and Goldman; S&R-H; Rabe-Hesketh et al.)
- "Fixed effects" versus "random effects" in TSCS data (e.g., Beck; Bafumi et al.; Wilson and Butler; Bafumi and Gelman; Western)
- Imputation methods for unbalanced panels or multilevel structures (e.g., S&R-H; Gelman et al.; Hsiao)
- Heterogeneity and state dependence in dynamic panel models (e.g., Heckman, Wooldridge; ask me about additional readings)
- Multilevel duration models (e.g., S&R-H; Sastry)
- Heterogeneity and duration dependence in multilevel duration models (Zorn; ask me about additional readings)
- Random effects multinomial logit models and the IIA assumption (S&R-H; Rudolph; ask me about additional readings)
- Models for heterogeneous event counts (e.g., Martin 2003)
- Cross-classified, non-nested models (e.g., Park and Jensen; S&R-H; R-H&S)