Quantitative Initiative for Policy and Social Research Advanced Multilevel Modeling Workshop

John Poe

May 2017

Description: This workshop is designed to (re)introduce you to basic hierarchical linear modeling from the perspectives of psychometrics (multilevel modeling), econometrics (correlated random effects modeling), and statistics (mixed effects modeling). We will then move beyond basic fixed and random effects modeling to more advanced topics for complex data structures and non-linear outcomes. Finally, we will walk through some of the conceptual and practical differences between standard (likelihood-based) multilevel modeling and Bayesian hierarchical modeling.

This workshop is not intended to be overly mathematical. It is designed to give you practical intuition behind complex model structure, diagnostics, and interpretation. By the end of the workshop you should (eventually) be able to do the following:

- Figure out how to set up basic and complex multilevel models in robust ways
- Interpret the right diagnostics to understand if you need more or less complicated models
- Understand the practical benefits of standard and Bayesian multilevel models
- Understand computational tradeoffs in speed, accuracy, the ability to produce any estimates at all (without changing software)
- Translate multilevel modeling papers written from an econometrics, psychometrics, and statistics perspective into a common language

Day 1: Thinking Through Multilevel Models

Morning 1: Overview and Basics of Clustered Data Problems
- Basic Assumptions and the Problems of Correlated Data
  - Standard Error Problems
  - Cluster Confounding
  - Heterogeneous Effects Across Groups
- The Semantics of Fixed, Random, and Mixed Effects Models Across Fields

Suggested Reading
➢ Gill, J. and A. J. Womack (2013). The Multilevel Model Framework. The SAGE handbook of multilevel modeling. M. A. Scott, J. S. Simonoff and B. D. Marx, Sage.

Morning 2: Multilevel Models
- Fixed and Random Effects
- Hierarchical Linear Modeling
- Group-Mean and Grand-Mean Centering in 2 or More Levels
- Random Coefficients Models

Suggested Readings
➢ Beck, N. and J. N. Katz (2007). "Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments." Political Analysis **15**(2): 182-195.
➢ Enders, C. K. (2013). Centering predictors and contextual effects. SAGE Handbook of Multilevel Modeling. M. A. Scott, J. S. Simonoff and B. D. Marx.

Afternoon 1: Some Implications of Badly Structured Data
- The Hausman Test & the Mundlak Test
- Too Few Groups
- Small Groups
- Correlated Groups
- Large Numbers of Levels
- Badly Modeled Subpopulations

Suggested Readings
➢ Ebbes, Peter, Ulf Böckenholt, and Michel Wedel. 2004. "Regressor and random-effects dependencies in multilevel models." *Statistica Neerlandica* 58 (2):161-78.
➢ Molenberghs, G. and G. Verbeke (2005). Non-Gaussian Random Effects. Models for discrete longitudinal data, Springer.
➢ Kim, Jee-Seon, and Edward W Frees. 2006. "Omitted variables in multilevel models." *Psychometrika* 71 (4):659-90.
➢ Raudenbush, Stephen W. 2008. "Many small groups." In *Handbook of multilevel analysis*: Springer.
➢ Li, Yi. 2009. "Modeling and Analysis of Spatially Correlated Data." In *New Developments in Biostatistics and Bioinformatics*.

- Stegmueller, D. (2013). "How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches." <u>American Journal of Political Science</u> **57**(3): 748-761.

Afternoon 2: Fit, Diagnostics, and Being Properly Paranoid about your Model
- The Intraclass Correlation Coefficient
- Wald tests, P values, & LR Tests
- Outlier Diagnostics
- Residual and Random Effect Diagnostics

Suggested Readings
- Snijders, Tom AB, and Johannes Berkhof. 2008. "Diagnostic checks for multilevel models." In *Handbook of multilevel analysis*: Springer.
- Van der Leeden, Rien, Erik Meijer, and Frank MTA Busing. 2008. "Resampling multilevel models." In *Handbook of multilevel analysis*: Springer
- Chambers, Raymond, and Hukum Chandra. 2013. "A random effect block bootstrap for clustered data." Journal of Computational and Graphical Statistics 22 (2):452-70.
    - lmeresampler: Bootstrap Methods for Nested Linear Mixed-Effects Model
- Loy, Adam, and Heike Hofmann. 2014. "HLMdiag: A suite of diagnostics for hierarchical linear models in R." *Journal of Statistical Software* 56 (5):1-28.
    - Loy, A. M. M. (2013). <u>Diagnostics for mixed/hierarchical linear models</u>. Dissertation

---

Day 2: Complex Structures

Morning 1: Multilevel Network Models
- Egocentric Network Models
- Actor-Partner Models
- Cross-Classified and Multiple Membership Models
- Whole Network Models

Suggested Readings
- Fielding, Antony, and Harvey Goldstein. 2006. "Cross-classified and multiple membership structures in multilevel models: An introduction and review."
- Rasbash, Jon, and William J. Browne. 2008. "Non-Hierarchical Multilevel Models." In *Handbook of multilevel analysis*, ed. J. de Leeuw and E. Meijer: Springer.
- Kenny, D.A. and Kashy, D.A., 2011. Dyadic data analysis using multilevel modeling. *Handbook of advanced multilevel analysis*, pp.344-360.
- Duijn, M. A. J. v. (2013). Multilevel Modeling of Social Network and Relational Data. *The SAGE handbook of multilevel modeling*. M. A. Scott, J. S. Simonoff and B. D. Marx, Sage.
- Mo, G. Y. and B. Wellman (2014). "Using Multiple Membership Multilevel Models to Examine Multilevel Networks in Networked Organizations."
- Tranmer, Mark, and Emmanuel Lazega. 2015. "Multilevel Models for Multilevel Network Dependencies." In *Multilevel Network Analysis for the Social Sciences*, ed. E. Lazega and T. A. Snijders: Springer.

➢ Snijders, T. A. (2016). The Multiple flavours of multilevel issues for networks. *Multilevel Network Analysis for the Social Sciences*: Springer

Morning 2: Multilevel Spatial Models
- Spatial Correlation at Different Levels
- Spatially Correlated Random Effects Modeling
- Spatial Random Coefficients

Suggested Readings
➢ Chaix, Basile, Juan Merlo, SV Subramanian, John Lynch, and Pierre Chauvin. 2005. "Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001." *American Journal of Epidemiology* 162 (2):171-82.
➢ Anselin, L., J. Le Gallo, et al. (2008). Spatial panel econometrics. The econometrics of panel data, Springer**:** 625-660.
➢ Li, Yi. 2009. "Modeling and Analysis of Spatially Correlated Data." In *New Developments in Biostatistics and Bioinformatics*.
➢ Baltagi, Badi H, Peter Egger, and Michael Pfaffermayr. 2013. "A generalized spatial panel data model with random effects." *Econometric Reviews* 32 (5-6):650-85.
➢ Blangiardo, Marta, and Michela Cameletti. 2015. *Spatial and spatio-temporal bayesian models with R-INLA*: John Wiley & Sons.
   o http://rpubs.com/INBOstats/spde
➢ Darmofal, David. 2015. *Spatial Analysis for the Social Sciences*, Cambridge University Press.
➢ Rousset, François. 2016. "An introduction to the spaMM package for mixed models." Unpublished Manuscript

Afternoon 1: Panel Data and Multilevel Modeling
- Fixed, Random and Mixed Effects Models
- Pattern Covariance Structures
- Lags

Suggested Readings
➢ Mundlak, Yair. 1978. "On the pooling of time series and cross section data." *Econometrica: Journal of the Econometric Society*:69-85.
➢ Nerlove, Marc. 2000. An essay on the history of panel data econometrics. Paper read at Proceedings of Ninth International Conference on Panel Data, Geneva, Switzerland.
➢ Fitzmaurice, Garrett M, Nan M Laird, and James H Ware. 2012. *Applied longitudinal analysis*. John Wiley & Sons. Ch. 7
➢ Bell, A. and K. Jones (2015). "Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data." Political Science Research and Methods **3**(01): 133-153
➢ Wang, L. P. and S. E. Maxwell (2015). "On disaggregating between-person and within-person effects with longitudinal data using multilevel models." Psychological Methods **20**(1): 63.

Afternoon 2: Repeated or Rolling Cross-Sections
- Fixed, Random, and Mixed Effects Models
- Growth Curve Models
- (Basic) Cointegration Problems

Suggested Readings
- Verbeek, M. (2008). Pseudo-Panels and Repeated Cross-Sections. <u>The Econometrics of Panel Data</u>.
- Choi, I. (2014). Panel Cointegration. The Oxford Handbook of Panel Data. B. H. Baltagi, Oxford University Press, USA.
- Lebo, M. J. and C. Weber (2015). "An Effective Approach to the Repeated Cross-Sectional Design." <u>American Journal of Political Science</u> **59**(1): 242-258.

Day 3: Generalized Linear Mixed Models

Morning 1: Likelihood, Simulated Likelihood, and Bayesian-flavored Multilevel Models
- Generalized Linear Mixed Models
- Optimization vs. Approximation vs. Simulation
- Link functions, priors, and distributional assumptions

Suggested Reading
- Browne, William J, and David Draper. 2006. "A comparison of Bayesian and likelihood-based methods for fitting multilevel models." *Bayesian analysis* 1 (3):473-514.
- Draper, David. 2008. "Bayesian multilevel analysis and MCMC." In *Handbook of multilevel analysis*: Springer.
- Stegmueller, D. 2013. "How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches." *American Journal of Political Science* 57 (3):748-61.

Morning 2: How Random Effects Are Estimated
- Linear Models
    - Maximum Likelihood (MLE)
    - Restricted Maximum Likelihood (REML)
- Non-Linear Models
    - Linearization
    - Laplace Approximation
    - Marginal Quasi Likelihood (MQL)/Penalized Quasi Likelihood (PQL)
    - Quadrature & Adaptive Quadrature
    - MCMC and Quasi-MCMC
    - Expectation Maximization (EM) and Variational Bayes

Suggested Readings
- Goldstein, Harvey, and Jon Rasbash. 1996. "Improved approximations for multilevel models with binary responses." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*:505-13.

- Bates, Douglas M, and José C Pinheiro. 1998. "Computational methods for multilevel modelling." *Madison, WI: University of Wisconsin*.
- Van der Leeden, Rien, Erik Meijer, and Frank MTA Busing. 2008. "Resampling multilevel models." In *Handbook of multilevel analysis*: Springer.
- Draper, David. 2008. "Bayesian multilevel analysis and MCMC." In *Handbook of multilevel analysis*: Springer.
- Train, Kenneth E. 2009. *Discrete choice methods with simulation*: Cambridge university press. Ch. 8-10
- Rabe-Hesketh, Sophia, and Anders Skrondal. 2012. "Multilevel and Longitudinal Modeling Using Stata." Ch. 10
- Fitzmaurice, Garrett M, Nan M Laird, and James H Ware. 2012. *Applied longitudinal analysis*. John Wiley & Sons. Ch. 15

Afternoon 1: Multilevel Binary Outcomes
- LPM, Probit, Logit, Clog-log, and Generalized Linear Mixed Models
- The Incidental Parameters Problem, Random Effects Misspecification, and other Nightmares
- What Likelihood, Simulated Likelihood, and Bayesian Variations Do Differently

Suggested Readings
- Beck, Nathaniel, Jonathan N Katz, and Richard Tucker. 1998. "Taking time seriously: Time-series-cross-section analysis with a binary dependent variable." *American Journal of Political Science* 42 (4):1260-88
- Lancaster, Tony. 2000. "The incidental parameter problem since 1948." *Journal of Econometrics* 95 (2):391-413.
- Greene, William. 2004. "The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects." *The Econometrics Journal* 7 (1):98-119.
- Crepon, Bruno, and Jacques Mairesse. 2008. "The Chamberlain Approach to Panel Data: An Overview and Some Simulations." In *The Econometrics of Panel Data*.
- Train, Kenneth E. 2009. *Discrete choice methods with simulation*: Cambridge university press.
- Huang, X. (2009). "Diagnosis of Random Effect Model Misspecification in Generalized Linear Mixed Models for Binary Response." *Biometrics* 65(2): 361-368.
- Arellano, M. and S. Bonhomme (2011). "Nonlinear Panel Data Analysis." *Annual Rev. Econ.* 3(1): 395-424.
- Rabe-Hesketh, Sophia, and Anders Skrondal. 2012. "Multilevel and Longitudinal Modeling Using Stata." Ch. 10
- Greene, W. (2014). Panel Data Models for Discrete Choice. The Oxford Handbook of Panel Data. B. H. Baltagi, Oxford University Press, USA.
- Beck, Nathaniel. 2015. Estimating grouped data models with a binary dependent variable and fixed effects: What are the issues? Paper read at annual meeting of the Society for Political Methodology, July.