Multilevel Models II

Assessment, estimation, generalization

Michele Claibourn, StatLab

June 27/28, 2013

Research Data Services in the Library

- Research Data Services: `www.library.virginia.edu/services/`
  - Data management plans
  - GIS training and consultations
  - Locating data, archiving data
- StatLab Services: `statlab.library.virginia.edu`
  - Individual consulting: advice, training or feedback on quantitative research
  - Workshops: Tentative fall schedule is available
- Upcoming Events

The Linear Mixed-Effects Model, Level 1

In scalar:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij} \tag{1}$$

Wtih $\varepsilon_{ij} \sim N(0, \sigma^2)$. In matrix form:

$$Y_j = X_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j$$

With $\varepsilon_j \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

The Linear Mixed-Effects Model, Level 2

In scalar form:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}z_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}z_j + u_{1j}\end{aligned} \tag{2}$$

With

$$E\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Var\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{bmatrix} = \mathbf{T}$$

In matrix form:

$$\boldsymbol{\beta}_j = Z_j \boldsymbol{\gamma} + \boldsymbol{u}_j$$

With $\mathbf{u}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{T})$

The Linear Mixed-Effects Model

By substitution, the linear mixed-effects model (in scalar) is:

$$y_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + u_{0j} + u_{1j} x_{ij} + \varepsilon_{ij} \tag{3}$$

In matrix:

$$Y_j = X_j Z_j \gamma + X_j u_j + \varepsilon_j$$

EDA

It's always a good idea to examine your data visually before embarking on model estimation.

- Looking at within-unit regression coefficients
  - Caterpillar plots: Is there variation to explain?
  - Scatterplots: What is the nature of the relationship between x and y?
- Looking at intercepts and slopes by level-2 covariates: What might explain variation across clusters?
  - Categorical
  - Continuous
- Assessing distributions: Are normal distributions a reasonable reference for inference?
  - Of level-1 ys and εs
  - Of random βs

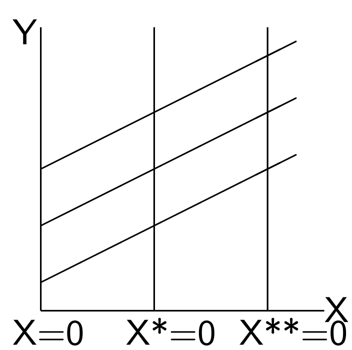
Model Building

Model specification should be informed by theory and data exploration.

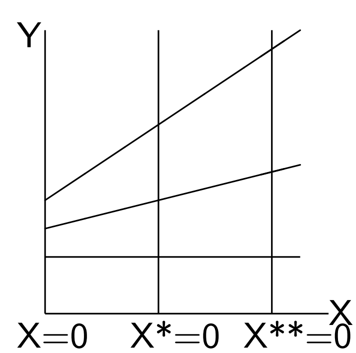
- The number of VCV parameters to be estimated increases rapidly with each random coefficient; more information is needed.
- Before deleting a level-1 coefficient, check for no slope heterogeneity as well as no average effect.
- With random intercepts and slopes, and interactions, give the *x*s meaningful zero points (via centering around the grand mean or group means)

Centering

Intercept variation



Intercept and slope variation



Estimation

Recall the general two-level model:

$$Y_j = X_j \boldsymbol{\beta}_j + \varepsilon_j$$

And

$$\boldsymbol{\beta}_j = Z_j \gamma + u_j$$

Combined:

$$Y_j = X_j Z_j \gamma + X_j u_j + \varepsilon_j$$

We must estimate γ, T and σ^2. From these we derive the estimates of β. We'll choose the estimates of the parameters that maximize the likelihood.

The Fixed Effects

The combined model for $\hat{\beta}_j$:

$$\hat{\boldsymbol{\beta}}_j = Z_j \gamma + u_j + e_j$$

The variance of $\hat{\beta}_j$, conditional on Z_j is:

$$var(\hat{\boldsymbol{\beta}}_j) = var(u_j + e_j) = T + \boldsymbol{\sigma}^2 (X_j' X_j)^{-1} = \boldsymbol{\Delta}_j$$

If the data are unbalanced, the Δ_j values will differ by cluster. If Δ_j is known, the unique, minimum-variance, unbiased estimator of γ would be the GLS estimator:

$$\hat{\gamma} = \left(\sum Z_j' \boldsymbol{\Delta}_j^{-1} Z_j \right) \sum Z_j' \boldsymbol{\Delta}_j^{-1} \hat{\beta}_j$$

The Likelihood Function

The likelihood function (generically):

$$L(\gamma, T, \sigma^2|y) \propto \prod_{j=1}^{J} p(y_j|\gamma, T, \sigma^2)$$

To get the marginal density of the data

$$L(\gamma, T, \sigma^2|y) = \prod_{j=1}^{J} \int p(y_j|u_j, \gamma, \sigma^2)p(u_j|T, \sigma^2)du_j$$

Where the conditional density of y, $p(y_j|u_j, \gamma, \sigma^2)$, is multivariate normal:

$$p(y_j|u_j, \gamma, \sigma^2) = \frac{exp[-(y_j - x_j z_j \gamma - x_j u_j)^2/2\sigma^2]}{(2\pi\sigma^2)^{\frac{n_j}{2}}}$$

And the marginal density of u, $p(u_j|T, \sigma^2)$, is multivariate normal:

$$p(u_j|T, \sigma^2) = \frac{exp((-u_j' T^{-1} u_j)/2)}{(2\pi)^{\frac{q}{2}}(T)^{\frac{1}{2}}}$$

Restricted Maximum Likelihood

REML adjusts for the uncertainty about the fixed effects; ML does not.

- In ML, estimates of variance and covariances are conditional on point estimates of the fixed effects (default in Stata)
- REML takes into account that these point estimates contain uncertainty and adjusts for this (default in R, SPSS, SAS, HLM)
- When differences occur, it is in the estimation of **T**; full ML variance estimates will be smaller, particularly when sample sizes are small.

REML: For any possible values of the fixed effects, γ, say γ_m, we can define a likelihood of T and σ^2: $L_m(T, \sigma^2|\gamma_m, y)$. Averaging over all possible values of $L_m(T, \sigma^2|\gamma_m, y)$ produces a likelihood of T and σ^2 given y alone.

The Random Coefficients

Two possibilities for predicting β_j:

1. The OLS regression coefficient based on data from each cluster:

$$\hat{\beta}_j = (X_j'X_j)^{-1}X_j'Y_j$$

2. The predicted value of β_j conditional on group characteristics in $\mathbf{Z}_j$:

$$\hat{\hat{\beta}}_j = Z_j\hat{\gamma}$$

The optimal, empirical Bayes (EB) combination of these:

$$\beta_j^* = \mathbf{\Lambda}_j\hat{\beta}_j + (I - \mathbf{\Lambda}_j)Z_j\hat{\gamma}$$

Where the weight, Λ_j is the ratio of the parameter variance matrix to the total variance matrix:

$$\mathbf{\Lambda}_j = T(T + \sigma^2(X_j'X_j)^{-1})^{-1}$$

Hypotheses Tests

Fixed effects

- Fixed parameters can be tested with a standard z-test.
- Multiple parameters (joint tests, linear constraints, etc.), can be tested via a Wald test.

Variance components

- LR test is preferred (for single and multiple parameters).
- To implement, estimate "full" model and model with some of the variance parameters set to zero.
- For models fit using REML, fixed effects must be the same across both models.
- For hypotheses on the boundary of the parameter space (i.e., 0), the significance level for the LR test is an upper bound.

In general, estimation problems tend to arise from estimation of the variance components.

Sample Sizes

Inference is based on asymptotics, so a key question is what sample size is large enough. Large enough for what?

- The power of the Wald test for level-2 fixed effects relies on the number of clusters; the power of the test for level-1 fixed effects relies on total sample size.
- In general, the standard errors of the variance components are more severely biased with small J.
- Some rules of thumb found in the literature (J/N): 30/30 or 10/30 for fixed effects; 50/20 or 30/20 for level-2 fixed effects; 100/10 or 50/10 for variance/covariance components.

The number of VCV components to be estimated is $\left[\frac{p(p+1)}{2}\right] + 1$ – where p is the number of random level-1 predictors in the model.

Model Assesment

Model Comparison

- Likelihood ratio tests
- Information criteria

Assessing Assumptions

- Distribution of level-1 errors
- Distribution of random effects

Shrinkage

More than two levels

The 3-level model may be written as:

$$y_{ijk} = \beta_{0jk} + \sum_{p=1}^{P} \beta_{pjk} x_{pijk} + \varepsilon_{ijk} \tag{4}$$

$$\beta_{pjk} = \gamma_{p0k} + \sum_{q=1}^{Q} Z_{qjk} + u_{pjk} \tag{5}$$

$$\gamma_{pqk} = \alpha_{pq0} + \sum_{s=1}^{S} \alpha_{pqs} w_{sk} + r_{pqk} \tag{6}$$

Where

- x_{pijk} =level-1 covariates: i indexes level-1 units, p indexes level-1 covariates
- z_{qjk} =level-2 covariates: j indexes level-2 units, q indexes level-2 covariates
- w_{sk} =level-3 covariates: k indexes level-3 units, s indexes level-3 covariates

The generalized linear mixed-effects model

The GLMM has a linear predictor component, i.e., the structural model, as well as a specific error structure depending on the type of nonlinear model estimated, and a link function, which governs how the nonlinear function is transformed.

GLM Families and Links You Already Know

Family	Link Function	Model
Gaussian	Identity	Linear model
Binomial	Logit	Logit model
Poisson	Log	Poisson count model

The GLMM for Binary Outcomes

1. Level-1 sampling model:
$$y_{ij} \sim Binom(\pi_{ij}, m_{ij}) \tag{7}$$
2. Level-1 link function, the logit:
$$\eta_{ij} = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) \tag{8}$$
The inverse of the link function:
$$\pi_{ij} = \frac{exp(\mathbf{x}\beta)}{1+exp(\mathbf{x}\beta)} \tag{9}$$
3. Level-1 structural model, same as above:
$$\eta_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \cdots + \beta_{pj}x_{pij} \tag{10}$$
4. Level-2 model, same as above:
$$\beta_{pj} = \gamma_{p0} + \sum_{q=1}^{Q} \gamma_{pq} z_{qj} + u_{pj} \tag{11}$$

Estimation of the GLMM

Obtaining ML estimates: (1) find the likelihood, this requires integration of the random effects from the joint distribution of the data and the random effects, (2) maximize the likelihood. Less easy when the model is nonlinear. Recall the joint distribution of the data and random effects

$$g(y, u|\gamma, T, \sigma^2) = p(y_j|u_j, \gamma, \sigma^2)p(u_j|T, \sigma^2) \tag{12}$$

And the likelihood of the data given the parameters

$$L(\gamma, T, \sigma^2, u|y) = \prod_{j=1}^{J} \int p(y_j|u_j, \gamma, \sigma^2)p(u_j|T, \sigma^2)du_j \tag{13}$$

NB: Produces joint distributions that are nonconjugate mixtures (can't be solved analytically). Integration achieved via numerical approximation – penalized quasi-likelihood, a Laplace approximation, (adaptive) Gauss-Hermite Quadrature... or via Bayesian algorithms.

Alternative Error Structures

Covariance of random effects

- Independent, exchangeable, identity, unstructured

Level-1 error

- Independent, exchangeable, ar, ma, etc.
- Heteroskedasticity

Useful References

- Bowers, Jake and Katherine Drake. 2005. "EDA for HLM: Visualization when Probabilistic Inference Fails." *Political Analysis* 13: 301-26.
- Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models,* 2nd Ed. Sage Publications.
- Snijders, Tom A.B. and Roel Boskers. 2012. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd Ed. Sage Publications.
- Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.
- Rabe-Hesketh, Sophia, and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata,* 3rd Ed. Stata Press.
- Pinheiro, Jose C. and Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-Plus.* Springer. (Refers to `nlme` package; for `lme4`, see Chapter drafts of *lme4: Mixed-effects Modeling with R*, available at `lme4.r-forge.r-project.org`)
- de Leeuw Jan and Erik Meijer, (ed.). 2008. *Handbook of Multilevel Analysis.* Springer-Verlag.