The Impact of Small Cluster Size on Multilevel Models: A Monte Carlo Examination of Two-Level Models with Binary and Continuous Predictors

Bethany A. Bell, Grant B. Morgan
University of South Carolina
Jeffrey D. Kromrey, John M. Ferron
University of South Florida
Introduction

- As the use of multilevel models has expanded into new areas, questions have emerged concerning how well these models work under various design conditions.
- Sample size at each level of analysis continues to be an important design condition in multilevel modeling.
Background

- Sample size ‘rules of thumb’ have been developed (e.g., 30 units at each level of analysis) for multilevel models.
- Many data sources in the social & behavioral sciences typically make these guidelines hard to achieve.
  - Complex sampling procedures often lead to large numbers of level-2 units with few individuals per cluster.
  - Evidence of the impact of level-2 sparseness with complex, “real-world” models is scarce.
This presentation includes findings from a study that focused on the consequences of level-2 sparseness on the estimation of fixed and random effects in terms of:

- model convergence rates
- statistical bias
- confidence interval accuracy and precision
- Type I error control
Monte Carlo Design

- **Level-1 Sample Size**
  - Small (average = 10, range 5 to 15)
  - Large (average = 50, range 25 to 75)

- **Level-2 Sample Size**
  - 50, 100, 200, 500

- **Proportion of Singletons**
  - 0, .10, .30, .50, .70

- **Levels of Collinearity**
  - 0, .30

- **Intraclass Correlation**
  - .05, .10, .15, .30

- **Model Complexity**
  - Continuous & binary predictors
  - K1 = 2, 3, 5
  - K2 = 1, 2, 4
  - Used in Nine Main Effect & Cross-Level Interaction Models
Model Specification

- After each data set was generated, the simulated sample was analyzed using a 2-level multilevel model with REML estimation and the Containment degrees of freedom estimation via the MIXED procedure in SAS.

- In all models, the intercept and level-1 coefficients were allowed to randomly vary and co-vary (i.e., an unstructured variance-covariance model specification)
Results: Convergence and Bias

- Model Convergence
  - More than 98% of the conditions evidenced no convergence problems
  - Highest rate of nonconvergence in the remaining 2% of conditions was less than 2% of the simulated samples

- Statistical Bias
  - Very low levels of statistical bias were evident for both fixed (min = -0.02, max = 0.02) and random (min = -0.01, max = 0.01) effects parameter estimates
Results: CI Coverage

- Overall, binary predictors at levels-1 and 2 behaved similarly to continuous predictors despite slightly larger CI widths
  - Proportion of singletons had no notable effect on the estimation of fixed effects for level-1 predictors
  - CI coverage for level-2 fixed effect parameters was reduced by proportion of singletons with smaller $N_2$ sample size
**Figure 1. Average coverage of level-2 predictors by level-2 sample size and proportion of singletons**

<table>
<thead>
<tr>
<th>Proportion of Singletons</th>
<th>Estimated Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1 (N=50)</td>
<td>0.88</td>
</tr>
<tr>
<td>W2 (N=50)</td>
<td>0.89</td>
</tr>
<tr>
<td>W3 (N=50)</td>
<td>0.90</td>
</tr>
<tr>
<td>W4 (N=50)</td>
<td>0.91</td>
</tr>
<tr>
<td>W1 (N=500)</td>
<td>0.92</td>
</tr>
<tr>
<td>W2 (N=500)</td>
<td>0.93</td>
</tr>
<tr>
<td>W3 (N=500)</td>
<td>0.94</td>
</tr>
<tr>
<td>W4 (N=500)</td>
<td>0.95</td>
</tr>
</tbody>
</table>

The graph shows the estimated coverage for different levels of singletons and sample sizes.
Figure 2. Bradley’s coverage of level-2 predictors by level-2 sample size and proportion of singletons
Results: Type I Error Rates

- Tended to be close to the nominal alpha of .05 across conditions for both fixed & random effects
  - Greatest departure from .05 was with the binary level-2 predictor
  - With large numbers of level-2 units ($N_2 = 500$), the proportion of singletons had limited effect on Type I error control of random effects
  - With fewer level-2 units ($N_2 = 50$), tests of random effects became conservative as the proportion of singletons increased
Figure 3. Average Type I error rate of binary level-2 predictor (W2) by level-2 sample size and proportion of singletons
Figure 4. Distribution of Type I error rates for tests of random effects by level-2 sample size and proportion of singletons
Discussion

- Researchers who have used sparse data structures to estimate multilevel models with binary or continuous predictors should not feel guilty.

- Proportion of singletons in the simulated samples had little impact on either the point or interval estimates of model parameters when large numbers of level-2 units were included.
Discussion

- With smaller level-2 sample sizes, increasing the proportion of singletons led to a reduction in the accuracy of the 95% CI for level-2 predictors but not for level-1 predictors.

- Model complexity, in terms of the number of predictors at each level and model type, did not impact our statistical outcomes.
Discussion

- Important to remember that findings are limited to the structure of the data and models included in this study
- Future studies include looking at dichotomous outcomes and linear models with violated assumptions