12-2017

Effect of Small Sample Size on Parameter estimation in a three-level Organizational Framework: A Simulation Study using M-Plus

Debrishi Bose
University of Connecticut, debrishi.bose@uconn.edu

Effect of Small Sample Size on Parameter estimation in a three-level Organizational Framework: A Simulation Study using M-Plus

Debrishi Bose

University of Connecticut

Spring 2017

Abstract

The use of Multilevel Modeling has become extremely popular in Social Science research owing to the natural hierarchy which often exists in the dataset. As has been pointed out by most researchers, number of units (sample size) at upper levels of hierarchy becomes extremely crucial. The current study employs a Random-Intercept model (using M-Plus) to study the effect of level-3 sample size on parameter estimation in a three level organizational framework. Number of sampling units at the third level was varied to check the impact on fixed effects, variance components, and their associated standard errors.

Keywords: Multilevel models, Small sample size, Relative Bias.

Effect of Small Sample Size on Parameter estimation in a three-level Organizational Framework: A Simulation Study using M-Plus

Introduction

In Educational Research, it is very common to encounter a dataset, which has a hierarchical structure, for example: the Office of Civil Rights (OCR) data and the Early Childhood Longitudinal Study (ECLS) data. The next question which often comes up is the number of units at each of the levels of the hierarchy, and if they're sufficient enough to draw valid conclusions. As highlighted in prior research like (Maas et al.,2005; McNeish et al.,2016), it is important to have more units at the higher levels (more clusters), however, the associated costs for increasing number of clusters poses a financial constraint on the part of researchers. Consequently, some simulation studies have been done by researchers to study the effect of small sample size on the estimation of parameters in a hierarchical framework. McNeish et al. (2016) refers to about twenty studies that has been conducted in the last decade or so, which focused to study the effects of sample sizes (both within and between clusters) on parameter estimation. If we carefully notice Table 1 (McNeish et al., 2016, pg. 300), we could easily verify that all the simulation studies conducted so far have only focused on two level hierarchical models with either continuous or binary outcomes. At the end of their simulation study, the authors also acknowledge the fact of the lack of simulation studies for three level hierarchical models along with other issues that still needs to be addressed in a multi-level framework. According to McNeish et al. (2016, pg. 311), “First, all research up to this point has focused on two-level models. Three-level models are relatively common in educational psychology (e.g., students clustered within classrooms/schools clustered within schools/districts), and sample sizes can become increasingly small as one progresses upward through a hierarchy. For instance, if school districts are the third level of clustering, even though five or ten school districts could provide data on thousands or even tens of thousands of students, the small sample size at the third level could lead to biased estimates". Thus, the main focus of this study would be to look at how the sample size affects estimation of parameters in a three-level hierarchical framework.

In his paper, McNeish et al. (2016) reviews those 20 studies and highlights the key elements- similarities and differences in those studies, thereby proposing guidelines or recommendations in terms of sample size for future research. I found another simulation study by Laszkiewicz (2013) along those lines. Although comparisons have been made between various studies, one should be careful while comparing the results since the results are essentially based on specific conditions that is inputted in the model like- the number of predictors at various levels, number of clusters, cluster size and even the ICCs. (Cools et al., 2007). So it is in the best interest to compare results from studies as long as the simulating conditions are similar.

Larger number of units are desirable at the group level in comparison to large number of units at the lower level. Lower number of units at the group level would reduce power to test for random slope variances across schools or clusters (Snijders, 2005, pg., 1570). While within-group sample size has a greater impact on level-1 estimates, number of clusters tend to influence level-2 estimates more (Harrow, 2002). The study that was conducted by (Maas et al., 2005) had one predictor at each of the two levels with three varying conditions-number of groups, number of individuals in each group and the ICC. The 27000 simulations that were carried out resulted in estimation procedure converging all the time producing admissible solutions. Inadmissibility was not a big concern that was reported by (Maas et al., 2005; Laszkiewicz, 2013; McNeish et al., 2016). Bell et al. (2008) reports that model converged 98% of the times out of the 5760 simulating conditions that was run during the study. As far as estimation of fixed effects are concerned, Maas et al. (2005) reports that both the fixed and random parameter estimates had a negligible bias (even less than 0.05%), with maximum bias occurring with the combination of lowest sample sizes and highest ICCs. Laszkiewicz (2013) found the average relative bias was found to be about 0.01% (almost unbiased), while it was somewhat higher (about 1.07%) for the random effects. McNeish et al. (2016) found that the number of clusters didn't have any impact on them, even with 5 clusters. However, Bell et al. (2008) found the estimated model parameters to be slightly negatively biased. Maas et al. (2005, pg., 89) reports the standard errors of variance components were marginally higher than the standard error associated with fixed effects, and the standard errors associated with group level variances were marginally underestimated. Laszkiewicz (2013) found that for a sample of size 25, variances in the intercept and slope were about 15% and 10% respectively, which decreased to less than 1% on increasing the numbers to groups to 100 or beyond. She also suggested a (10 clusters/5 cluster size) rule for unbiased estimation, which is way different from the (30 Clusters/30 Cluster-size) rule as suggested by Kreft (1996). In the simulation study of McNeish et al. (2016), the authors used a simple balanced model for illustrative purposes with a continuous outcome and one level-1 predictor. The level-1 variances were almost unaffected, with maximum "percentage underestimated" being

0.30%. Level2 variances and associated standard errors were showed to exhibit a pattern when the number of clusters were less than 30.

Methods:

In three level MLM, level 1 coefficients are treated as outcomes of level 2 equations, and level 2 coefficients are outcomes of the level 3 equations, and hence the outcome variable at the lowest level is being modeled by predictors from all the levels (Subedi et al, 2005). A Monte Carlo simulation study was conducted using MPlus Version 8. The outcome variable was Reading Achievement Score. For simplicity purposes, a Random Intercept Model was used with two predictors at level 1. With regard to the population parameters of this study, Model 2 of Section 3 of the paper Bell et al. (2013) was used to generate the data. Following the notation of three level models as outlined in Subedi (2005, pg., 32), the model used for this case is defined below:

$$\boldsymbol{Y_{ijk}} = \boldsymbol{\pi_{0jk}} + \boldsymbol{\pi_{1jk}} * \boldsymbol{X_{1ijk}} + \boldsymbol{\pi_{2jk}} * \boldsymbol{X_{2ijk}} + \boldsymbol{e_{ijk}}.$$

$$\boldsymbol{\pi_{0jk}} = \boldsymbol{\beta_{00k}} + \boldsymbol{r_{0jk}},\ \boldsymbol{\pi_{1jk}} = \boldsymbol{\beta_{10k}} + \boldsymbol{r_{1jk}},\ \boldsymbol{\pi_{2jk}} = \boldsymbol{\beta_{20k}} + \boldsymbol{r_{2jk}}.$$

$$\boldsymbol{\beta_{00k}} = \boldsymbol{\gamma_{000}} + \boldsymbol{u_{ook}},\ \boldsymbol{\beta_{10k}} = \boldsymbol{\gamma_{100}} + \boldsymbol{u_{10k}}, \boldsymbol{\beta_{20k}} = \boldsymbol{\gamma_{200}} + \boldsymbol{u_{20k}}.$$

$where,\ \boldsymbol{Y_{ijk}}\ denotes\ the\ achievement\ score\ of\ student$ "i" $in\ classroom$ j" for the school k",

$\boldsymbol{X}'s\ denote\ the\ predictors\colon\ \boldsymbol{\pi}, \boldsymbol{\beta}\ \boldsymbol{and}\ \boldsymbol{\gamma}\ denotes\ fixed\ effects, \boldsymbol{e}/\boldsymbol{r}\ denotes\ the\ residuals$

Since we are mainly interested in studying the impact of small number of level 3 units on parameter estimation, the number of level 3 units were varied from 10-100 at intervals of 10. The number of level 1 units (20), level 2 (5) and ICCs were kept fixed throughout the study. Maximum Likelihood procedure was used to estimate the models.

Results & Discussion Model

Convergence:

With regard to the simulation study results, these not are very comparable to the previous ones since those were conducted in a two level framework while this one employs a three level hierarchical framework. Most of the previous researchers did not report any major issues involving model convergence (already discussed in the paper), however, in this case, issues involving model convergence have been seen. Out of the 500 replications requested for each of the 10 conditions, all the cases produced less than 90% of the replications. The least number of replications (409) was produced corresponding to N=20, while the maximum number (437) was produced for cases corresponding to N=70. In most cases, 420-430 replications were produced. Overall, the percentage of replications varied between 82% and 88%. Even though a random Intercept only model was used, but it resulted in model Non convergence 12-18% of the times, perhaps more number of units are needed at the 3rd level.

Fixed Effects:

Table 1: Relative Average Bias Percentage for Fixed Effects

Number of level-3 units	LEP	EC
100	14	-0.08
90	14	-0.02
80	14	1.11
70	42	0.88
60	0	0.44
50	28	0.66
40	28	1.11
30	28	1.11
20	43	1.55
10	71	0.22

From the table presented above, it is clearly evident that the percentage of Bias for the two fixed effects look completely different. The bias percentages differed in two aspects- magnitude and direction (over or under estimation). For the first one, bias percentages were calculated (as outlined in Muthen et al., 2002, pg., 8) and were to be positive, indicating that it was over estimated (according to the notation of (McNeish et al., 2016, pg., 300)) in all the cases, except one, where bias percentage was 0, corresponding to sample size =60. It is clear that for smaller sample size, it was as high as 71% corresponding to 10 level 3 units. For the second fixed effect, the bias percentage was marginal for all the cases. However, for sample size = 90 & 100, it is slightly underestimated, while for other cases, they were slightly over estimated. They varied from -0.02% to 1.55%, which is considerably smaller in comparison to the bias percentages of other fixed effect. Hence, relative bias percentages for one of the level 1 fixed effects are consistent with previous studies, while the other one is not.

Table 2: Relative Average Bias Percentage for Standard Error of Fixed Effects

Number of level-3 units	LEP	EC
100	-27	2066
90	3	1166
80	32	4300
70	-47	4150
60	-42	2325
50	-13	12825
40	-68	19900
30	-6	9500
20	-40	2328
10	40	300

With regard to the standard errors associated with both the fixed effects, they are significantly higher as compared to what has been reported in previous research. Even the bias percentages in the standard errors of the two fixed effects are very much dissimilar. While in one case, they varied from -68% to 40%, and in the other, it fluctuated between 300% to about 20000%. Hence, it is clearly evident that small number of units have a tremendous impact on the standard errors associated with fixed effect estimates in a three-level framework.

Variance Estimates:

Table 3: Average Relative Bias for Variance Components

Number of level-3 units	Level1	Level2	Level3
100	-30	125	453
90	-64	158	174
80	17	116	189
70	-113	125	153
60	-30	175	295
50	-15	125	247
40	-22	150	263
30	83	133	379
20	-12	116	742
10	12	108	858

From the table presented above, it is clearly evident that the estimates of variance at all the levels had significant bias. This average relative bias is relatively much higher than the average bias involved with fixed effects estimation. Previous research reported by (Maas et al., 2005; Laszkiewicz, 2013; McNeish, 2016) also showed variance components to be more affected by smaller sample size. However, in this case, the associated bias percentages are significantly greater than those reported from two level studies. While level-1 variance estimates have been both over and under estimated, the variance estimates at the higher levels have consistently been over estimated. Furthermore, it is to be noted that the magnitude of average relative bias increases as one moves up the hierarchy. The average relative bias varied from -113%-83%, 100%-175% and 153%-858% for level 1, level 2 and level 3 variance components respectively. Thus, variance estimates are severely biased due to the small sample size.

Table.4: Avg Relative Bias for Std.Err in Var Component

Number of level-3 units	Level1	Level2	Level3
100	60	7024	270
90	-57	993	1152
80	457	3094	2109
70	-12	2909	2347
60	356	6451	4030
50	1743	2581	572
40	2414	785	702
30	565	1287	333
20	2178	8776	1693
10	223	16089	294

The relative bias in the standard errors of variance estimates are even worse and extremely fluctuating across all the conditions. As originally proposed by Raudenbush and Bryk (2002) and highlighted by McNeish (2016), "standard errors associated with level 2 variance components are 4-th order estimators, and hence they require a good amount of data to be properly estimated". Furthermore, in this case, level 3 variance components have been estimated. Following the same argument, it is not very surprising that the magnitude of standard errors associated with variance components are so high. Perhaps, more than 100 clusters are needed for unbiased estimation of variance and associated standard errors in a 3level HLM.

Conclusion

Since this was a first attempt to conduct a simulation study in a three-level organizational framework to check the impact of small sample size on parameter estimates, a relatively easy model was chosen. Models which are most likely to be used in such situations would perhaps be more complex. Unlike other studies, the number of varying conditions were also less. Furthermore, due to the absence of three-level simulation studies in Educational field, results obtained from this study is not also comparable. Thus, there is certainly a need to conduct more such studies to check for consistency in the results. Future studies could be set up using a more complex model (by incorporating random slopes or more predictors) with more varying conditions. According to Kanten et al (2015), learning organizations and organizational structure have a significant impact on individuals and groups in terms of the outputs they produce. So in educational research, it is important to study the impact of such predictors (at various levels) on the achievement scores of students by employing MLMs. These studies would also help organizations to be aware of their strengths or weaknesses (in terms of what works and what does not), and hence appropriate decisions in terms of framing of its policies could be taken.

References:

Bell, B.A, Ferron, J.M., & Kromrey, J.D. (2008). Cluster Size in Multilevel Models: The Impact of Sparse Data Structures on Point and Interval Estimates in Two-level Models. *Section on survey Research Methods- JSM.*

Bell, B.A, Ene, M., Smiley, W, Schoeneberger, J.A (2013). A Multi-level Model Primer using SAS PROC MIXED. *Statistics and Data Analysis*. pp 433-452.

Cools,W. & Van den Noortgate,W (2007). Power and Accuracy in Multilevel Designs: An application in Educational Research. *International Statistical Institute*, 56th Session.

Harrow, K. H. (2002). A Comparison of Three Estimation Methods for Hierarchical Linear Models: A Simulation Study (Doctoral Dissertation). UMI Number: 3067333

Kanten, P, Kanten, S., & Gurlek., M (2015). The Effects of Organizational Structures and Learning Organization on Job Embeddedness and Individual Adaptive Performance. *Procedia Economics and Finance*, Vol 23, pp 1358-1366.

Laszkiewicz, E. (2013). Sample size and Structure for Multilevel Modelling: Monte Carlo Investigation for the Balanced Design. *Quantitative Method in Economics*, Vol. 14, No.2, pp. 19-28.

Maas, C.J.M & Hox, J.J (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology*. Vol. 1(3), pp. 86-92, DOI 10.1027/1614-1881.1.3.86

McNeish, D.M, Stapleton, L.M. (2016). The Effect of Small Sample Size on Two Level Model Estimates: A Review and Illustration. *Educational Psychology Review*. Volume 28, Issue 2, pp 295–314.

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén

Muthén, L.K. and Muthén, B.O. (2002). How to Use a Monte Carlo study to decide on Sample Size and determine Power. *Structural Equating Modeling*. Vol. 9, Issue. 4, pp. 599-620.

Snijders, T.A.B. (2005). Power and Sample size in Multilevel Modeling. *Encyclopedia of Statistics in Behavioral Science.* Vol. 3, pp. 1570-1573.

Subedi, B.R. (2005). A Demonstration of the Three-level Hierarchical Generalized Linear Model Applied to Educational Research. *Florida State University Libraries*.