Journal of Biomedical Analytics
Vol. 1, No. 2 (2018), pp. 81–89

doi:10.30577/jba.2018.v1n2.26

Research Article

Effect of Sample Size on the Profile Likelihood Estimates for Two-stage Hierarchical Linear Models

Noor Jahan Akter[1] and Md Hasinur Rahaman Khan[2,*]

[1,2]Institute of Statistical Research and Training, University of Dhaka, Bangladesh
[2]Corresponding author: hasinur@isrt.ac.bd

Received: Aug 31, 2018; revised: Sep 11, 2018; accepted: Oct 12, 2018.

Abstract: Determining sample size to produce accurate parameter estimates and make valid inferences about population parameter is a pivotal problem in any well planned research. The issue of sample size becomes more complex in hierarchical linear models because the units at different levels are hierarchically nested. Many studies have been carried out to determine sample sizes at different levels of a nested data model. However, most of the studies assume that the sample size is large enough to conduct the significance test. Some alternative methods can be used to relax the assumption of large sample. *Profile likelihood method* is more robust in case of small sample and when the variance components need to be estimated. In this study, we investigate the effect of sample sizes on the performance of parameter estimates at the group-level and individual-level of a two-level regression model. We consider a more appropriate statistical approach, profile likelihood method, to check the reliability of estimates of fixed coefficients and variance components.

Keywords: multilevel analysis, hierarchical linear model, sample size, maximum likelihood, profile likelihood interval

1 Introduction

Multilevel analysis emerged in sociological studies. With the development of its theory, researchers of diverse fields have started to consider the technique where units of analysis are hierarchically nested (Khan and Shaw, 2011). Recently, multilevel analysis has become a buzz word in public health research. The basic model of multilevel analysis is the *Hierarchical Linear Model*(HLM) where the hierarchies in the data are reflected by the distribution of the errors in the model. Traditional statistical techniques assume that errors are independent and identically distributed. HLM takes into account the dependency of the observations within groups by assigning higher level variables at the group-levels. The interaction effects between individual-level and group-level variables can also be assessed by HLM.

Determining sample size to produce accurate parameter estimates and make valid inferences about population parameter is a pivotal problem in any well planned research. The issue of sample size becomes more complex in hierarchical linear models because the units at different levels are hierarchically nested. A rule of thumb, called the '30/30' rule, is recommended by Kreft (1996). Later, Hox (1998) proposed a minimum of 20 observations at level-1 and 50 groups at level-2 when examining cross-level interaction across levels. Maas and Hox (2004, 2005) found that when the number of groups is less than 100, group-level standard errors are biasedly estimated. Clarke and Wheaton (2007) suggested that for accurate parameter estimates of level-1 residual variance at least 100 groups and 5 observations per group is required. Reliable parameter estimates for intercept variance can be found when there are at least 100 groups with 10 observations per each group. For approaching the parameter value of group-level slope variance, they advised to have a group size 20 and number of groups at least 200.

Most of the simulation studies investigated the effect of sample size based on the asymptotic theory, that is, they assumed that the sample size is sufficiently large to conduct the significance test. Even though the sample size is large enough, testing a variance component using their standard error is woefully inadequate (Bates, 2010). Van der Leeden *et al.* (1997) found that the estimates of standard errors for variance components are usually underestimated. To test variance components, Berkhof and Snijders (2001) got modified likelihood-ratio test better by means of a Monte Carlo simulation study. For testing the fixed model terms of linear mixed models when the sample size is small, Welham and Thompson (1997) proposed Bartlett-type adjustments in likelihood ratio tests. Manor and Zucker (2004) compared several methods for evaluating their performance in small sample to test the fixed effects in the mixed linear model.

When the issues are small sample and estimating variance components, *profile likelihood method* is expected to be more robust. When the distribution is skewed, Wald-type intervals fail to estimate the uncertainty of variances. In such circumstance, profile likelihood interval performs better by allowing uneven distances of the lower and upper confidence limits to the point estimates (Gerhard, 2010). In this study, we investigate the effect of sample sizes on the performance of parameter estimates at the group-level and individual-level of a two-level regression model. We consider a more appropriate statistical approach, profile likelihood method, to check the reliability of estimates of fixed coefficients and variance components.

2 Method

We consider two-level regression model with one regressor at each level.

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}, \\ \beta_{0j} &= \gamma_{00} + \gamma_{01}Z_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_j + u_{1j}, \end{aligned} \tag{1}$$

where

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u0}^2 & \sigma_{01} \\ \sigma_{01} & \sigma_{u1}^2 \end{bmatrix}\right).$$

Here Y_{ij} is the outcome variable for i^{th} individual in j^{th} group, X_{ij} is the level-one regressor, and Z_j is the level-2 regressor. Model (1) can be rewritten as

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{0j} + u_{1j}X_{ij} + e_{ij}. \tag{2}$$

The intra-class correlation coefficient (ICC) is another factor that may have influence on the accuracy of the parameter estimates (Goldstein, 2011). In the hierarchical linear model, the ICC ρ is defined as

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}, \tag{3}$$

where σ_e^2 is the lowest-level residual variance and σ_{u0}^2 is the random intercept variance in a fully unconditional hierarchical linear model $Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$, where $u_{0j} \sim N(0, \sigma_{u0}^2)$. In words, Equation 3 is the proportion of the variance explained by the clustering structure in the population (Hox *et al.*, 2017).

Following the simulation scenarios used by Maas and Hox (2004, 2005), we will look at three conditions: (1) Number of groups (4 conditions: 20, 30, 40 and 50), (2) Group Size (4 conditions: 5, 15, 25 and 30) and (3) Intraclass Correlation (4 conditions: 0.1, 0.2, 0.3, and 0.4). The level-1 and level-2 regressors X and Z are distributed as standard normal. The lowest level variance σ_e^2 is set to 0.5 (Maas and Hox, 2004, 2005). The group random components u_{0j} and u_{1j} are independent normal variables with mean zero and variance σ_{u0}^2 and σ_{u1}^2 (Maas and Hox, 2004). The residual variance σ_{u0}^2 at group-level follows from the ICC and σ_e^2 (Maas and Hox, 2004, 2005). The effect for the intercept variance and the slope variance are similar (Busing, 1993) , as a result σ_{u0}^2 and σ_{u1}^2 can be set to equal (Maas and Hox, 2005). We set the fixed effect parameters $\gamma_{00} = 1$, $\gamma_{10} = 0.3$, $\gamma_{01} = 0.3$, $\gamma_{11} = 0.3$ (Cohen, 1988; Maas and Hox, 2005). The covariance between the intercept variance and the slope variance σ_{01} is set to zero (Maas and Hox, 2004, 2005). The random error terms e_{ij}, u_{0j} and u_{1j} are then generated after setting the true values of the parameters.

Full maximum likelihood (FML) and Restricted maximum likelihood (REML) are two most common variants of estimation techniques in multilevel models. REML is used in this simulation study to estimate the regression coefficients and variance components. The reason behind choosing REML over FML is REML provides better estimates than FML (Browne, 1998). To check the unbiasedness of the parameter estimates, we calculate percentage relative bias which is defined by $\frac{\hat{\theta}-\theta}{\theta} \times 100$, for the estimate $\hat{\theta}$ of the population parameter θ. 95% profile likelihood interval is used to examine the reliability of fixed effect and random effect parameters.

There are $4 \times 4 \times 4 = 64$ simulated combinations and we generated 1000 Monte Carlo data sets for each combination. We used software `R` for simulating observations and estimating parameters. The restricted maximum likelihood estimates of the parameters in the hierarchical linear model were determined using the `lmer` function of the `lme4` package (1.1-13) for `R` version 3.4.1.

3 Analysis

Table 1 shows the percentage relative bias of estimates for different number of groups, group size and ICC. We found the relative biases for the standard error of lowest-level residual variance and fixed effect parameter estimates are unimportant for every simulated condition. The biases in the estimates of standard errors of group-level variances are comparatively bigger than the fixed effect parameter estimates. Table 1 reveals that estimates of standard errors of intercept and slope variances are prone to be biased downward. The largest biases for σ_{u0} and σ_{u1} are -11.446% and -13.502% respectively when there are 20 groups with 5 observations per each group. However, the biases for the estimates of standard errors of group-level variances are less than 3.5% for number of groups 30 and group size 15. The biases for ICC=0.1 are relatively higher than the other values of ICC.

Table 1: Relative bias of estimates for different number of groups, group size, and ICC.

Number of groups	Group size	ICC	σ_{u0}	σ_{u1}	σ_e	γ_{00}	γ_{10}	γ_{01}	γ_{11}
20	5	0.1	-11.446	-13.502	-0.993	0.095	-2.555	0.818	0.355
		0.2	-6.096	-7.773	-0.510	0.118	-3.156	1.016	0.442
		0.3	-2.747	-3.883	-0.593	0.430	-4.925	-0.518	1.334
		0.4	-3.005	-3.129	-0.340	-0.869	-1.715	-2.283	2.134
	15	0.1	-5.208	-4.352	0.181	-0.150	-0.211	-0.345	0.884
		0.2	-3.512	-1.248	-0.102	0.145	0.497	1.179	0.247
		0.3	-0.541	-2.765	0.159	0.029	-2.340	-1.025	-1.625
		0.4	-1.537	-2.698	-0.064	-0.141	1.266	-0.338	1.280
	25	0.1	-3.422	-2.679	0.003	-0.077	-0.905	-0.362	0.540
		0.2	-2.487	-3.625	0.167	-0.264	1.548	1.765	0.362
		0.3	-2.261	-2.301	0.009	-0.228	-0.280	-0.851	-2.954
		0.4	-1.473	-1.075	-0.208	-0.001	0.530	0.329	1.365
	30	0.1	-2.831	-1.989	-0.076	-0.260	0.877	-1.295	2.230
		0.2	-2.307	-1.004	-0.214	-0.156	-0.160	-1.211	-1.833
		0.3	-0.643	-1.097	-0.099	-0.381	0.156	-0.420	2.090
		0.4	-1.929	-2.136	-0.059	-0.787	-1.992	1.088	0.694
30	5	0.1	-11.393	-8.145	-0.434	0.198	0.034	1.011	0.065
		0.2	-4.391	-4.018	-0.459	-0.302	0.110	0.369	-0.830
		0.3	-3.016	-2.798	-0.060	-0.058	-0.285	-1.318	-1.774
		0.4	-1.750	-2.011	-0.270	0.210	0.156	-1.804	-0.219
	15	0.1	-1.926	-3.178	0.153	-0.114	0.962	0.024	-0.806
		0.2	-1.672	-1.856	-0.024	-0.205	-0.618	-0.729	0.562
		0.3	-1.666	-1.816	-0.256	0.186	0.033	0.455	0.760
		0.4	-0.021	-1.196	0.195	-0.405	1.081	-0.940	-1.647
	25	0.1	-2.509	-2.945	-0.174	0.275	0.474	-0.064	0.330
		0.2	-1.418	-0.950	-0.159	0.128	0.931	-0.715	-0.425
		0.3	-0.986	-1.348	-0.009	0.142	-1.516	1.930	-0.682
		0.4	-0.594	-1.819	0.102	-0.008	2.404	-0.739	1.144
	30	0.1	-1.580	-1.018	0.037	0.003	-0.527	0.342	-0.131
		0.2	-1.230	-1.582	0.034	-0.265	-1.488	-0.430	1.180
		0.3	-0.629	-0.516	-0.198	0.051	-0.135	-0.089	-0.940
		0.4	-1.799	-0.147	-0.059	-0.061	-1.074	-1.610	1.522
40	5	0.1	-7.136	-6.709	-0.491	0.152	0.215	-1.466	0.147
		0.2	-2.510	-3.855	-0.359	0.232	-0.169	1.268	3.039
		0.3	-0.917	-1.726	-0.116	-0.239	0.704	-0.919	-0.399
		0.4	-0.704	-1.278	0.018	-0.145	0.071	-0.948	0.520
	15	0.1	-1.915	-1.062	-0.001	0.155	-0.104	-1.030	1.097
		0.2	-0.673	-1.055	-0.022	0.296	-0.142	1.510	-0.617
		0.3	-1.082	-0.982	0.026	-0.072	0.782	-0.775	0.529
		0.4	-0.134	-0.639	0.152	0.071	-2.585	-0.856	-0.041
	25	0.1	-0.072	-1.126	-0.126	-0.045	0.286	-0.408	0.425
		0.2	-0.904	-1.315	0.009	-0.113	-0.254	0.076	-0.208
		0.3	-0.608	-0.575	-0.049	0.108	-1.233	0.660	-0.114
		0.4	-1.083	-1.130	-0.111	0.195	-0.908	-1.817	-2.725
	30	0.1	-1.886	-1.662	-0.087	-0.257	-0.014	0.081	0.610
		0.2	-1.380	-0.908	-0.094	-0.185	2.049	0.473	1.593
		0.3	-0.302	-1.278	0.126	0.548	-0.444	-0.141	0.166
		0.4	-1.052	-0.981	0.012	0.331	-0.560	-0.460	-1.203

50	5	0.1	-6.435	-6.613	-0.535	0.188	0.085	0.396	0.550
		0.2	-2.740	-1.904	-0.411	0.245	-0.070	0.301	0.076
		0.3	-1.345	-2.396	-0.260	0.234	0.283	0.277	0.700
		0.4	-1.274	-1.294	-0.258	0.533	-0.968	-1.301	0.082
	15	0.1	-1.823	-2.690	-0.102	0.103	0.031	0.412	-0.466
		0.2	-1.530	-0.614	0.009	-0.061	-0.157	0.756	-0.249
		0.3	-1.005	-1.041	-0.027	0.330	0.374	0.411	0.857
		0.4	0.066	-1.323	0.000	-0.206	0.658	0.453	-0.345
	25	0.1	-0.717	-1.048	-0.026	-0.005	-0.228	-0.459	0.194
		0.2	-0.758	-0.874	-0.069	0.054	0.509	0.452	-1.045
		0.3	-0.578	-0.979	0.076	-0.106	1.479	-0.039	-0.651
		0.4	-0.561	-1.250	0.076	-0.519	0.733	0.172	0.764
	30	0.1	-1.403	-1.161	-0.158	0.142	-0.071	0.074	-0.515
		0.2	-1.125	-0.742	-0.077	-0.111	0.151	-0.086	0.053
		0.3	-0.763	-0.647	0.026	-0.247	1.341	0.207	0.503
		0.4	-0.482	-0.877	0.024	0.280	-0.659	0.618	0.408

Table 2, Table 3 and Table 4 display the non-coverage probabilities of 95% profile likelihood interval by number of groups, group size and ICC, respectively. Note that, we computed non-coverage rates on the standard deviation scale of variance components. From Table 2, we see that non-coverage rates for fixed effect parameters range from 5.6% to 6.9%. For the standard errors of group-level variances, non-coverage rates are higher for number of groups 20 where the non-coverage rate is 7.8% for σ_{u0} and 7.9% for σ_{u1}. The non-coverage rate for the standard error of lowest-level variance is close to nominal non-coverage rate 5%. Overall, as the number of groups increases, the non-coverage probabilities of 95% profile likelihood interval for both the fixed effect and random effect parameters approach to the nominal 5%.

From Table 3, we see that coverage rates for fixed effect parameters are better than group-level variance components. But we did not find any trend of non-coverage rate for group size. One possible reason behind it can be we have chosen nearer group size in our simulation. Also, Table 4 shows no pattern of non-coverage rate for the change in ICC. Table 5 shows the rates of non-coverage for each parameter in each simulated data set. The non-coverage rates for the fixed effect parameters range from 4.4% to 8.4% and for the level-2 variance components, it is between 4.3% to 10.5%.

Table 2: Non-coverage of the 95% Profile Likelihood Interval by Number of Groups.

Parameter	Number of groups			
	20	30	40	50
σ_{u0}	0.078	0.068	0.061	0.060
σ_{u1}	0.079	0.070	0.068	0.062
σ_e	0.052	0.047	0.051	0.050
γ_{00}	0.069	0.061	0.058	0.059
γ_{10}	0.068	0.062	0.062	0.059
γ_{01}	0.065	0.064	0.056	0.056
γ_{11}	0.069	0.062	0.058	0.057

Table 3: Non-coverage of the 95% Profile Likelihood Interval by Group Size.

Parameter	Group Size			
	5	15	25	30
σ_{u0}	0.063	0.065	0.069	0.070
σ_{u1}	0.070	0.069	0.071	0.069
σ_e	0.050	0.050	0.050	0.051
γ_{00}	0.062	0.065	0.061	0.060
γ_{10}	0.061	0.061	0.062	0.066
γ_{01}	0.060	0.062	0.061	0.059
γ_{11}	0.062	0.062	0.061	0.061

Table 4: Non-coverage of the 95% Profile Likelihood Interval by Intra-class Correlation.

Parameter	Intra-class Correlation			
	0.1	0.2	0.3	0.4
σ_{u0}	0.066	0.065	0.067	0.069
σ_{u1}	0.068	0.070	0.070	0.071
σ_e	0.046	0.053	0.052	0.050
γ_{00}	0.060	0.064	0.060	0.063
γ_{10}	0.064	0.062	0.064	0.061
γ_{01}	0.059	0.061	0.064	0.060
γ_{11}	0.064	0.060	0.061	0.061

Table 5: The effect of number of groups, group size, and ICC on the non-coverage of the 95% profile likelihood interval.

Number of groups	Group size	ICC	σ_{u0}	σ_{u1}	σ_e	γ_{00}	γ_{10}	γ_{01}	γ_{11}
20	5	0.1	0.049	0.043	0.052	0.068	0.066	0.065	0.076
		0.2	0.068	0.089	0.052	0.068	0.062	0.066	0.075
		0.3	0.066	0.082	0.054	0.067	0.067	0.063	0.076
		0.4	0.083	0.08	0.045	0.068	0.069	0.067	0.05
	15	0.1	0.086	0.081	0.056	0.065	0.069	0.073	0.084
		0.2	0.076	0.071	0.054	0.076	0.057	0.068	0.062
		0.3	0.086	0.085	0.055	0.063	0.068	0.073	0.074
		0.4	0.063	0.083	0.048	0.076	0.069	0.058	0.064
	25	0.1	0.086	0.081	0.042	0.068	0.069	0.067	0.068
		0.2	0.079	0.103	0.058	0.067	0.061	0.07	0.072
		0.3	0.076	0.07	0.056	0.066	0.077	0.07	0.071
		0.4	0.083	0.076	0.041	0.079	0.062	0.06	0.06
	30	0.1	0.105	0.085	0.044	0.067	0.079	0.061	0.079
		0.2	0.085	0.071	0.065	0.067	0.078	0.059	0.061
		0.3	0.074	0.076	0.052	0.063	0.071	0.058	0.066
		0.4	0.081	0.082	0.062	0.076	0.066	0.067	0.073
30	5	0.1	0.055	0.064	0.049	0.055	0.056	0.067	0.055
		0.2	0.069	0.066	0.047	0.062	0.052	0.072	0.075
		0.3	0.079	0.082	0.053	0.056	0.073	0.053	0.062
		0.4	0.068	0.068	0.051	0.063	0.056	0.067	0.06
	15	0.1	0.056	0.073	0.041	0.074	0.065	0.052	0.066
		0.2	0.052	0.065	0.053	0.066	0.065	0.064	0.048
		0.3	0.064	0.071	0.046	0.072	0.062	0.07	0.067
		0.4	0.061	0.069	0.046	0.049	0.06	0.072	0.073
	25	0.1	0.071	0.079	0.03	0.054	0.064	0.059	0.054
		0.2	0.076	0.07	0.055	0.061	0.067	0.064	0.061
		0.3	0.068	0.059	0.045	0.058	0.049	0.054	0.056
		0.4	0.078	0.073	0.059	0.058	0.059	0.064	0.071
	30	0.1	0.072	0.08	0.047	0.061	0.073	0.063	0.066
		0.2	0.063	0.066	0.045	0.063	0.069	0.069	0.058
		0.3	0.075	0.064	0.043	0.066	0.066	0.063	0.063
		0.4	0.074	0.067	0.041	0.056	0.06	0.063	0.049

40	5	0.1	0.06	0.062	0.051	0.054	0.062	0.05	0.056
		0.2	0.054	0.08	0.056	0.048	0.066	0.058	0.058
		0.3	0.053	0.068	0.047	0.052	0.056	0.059	0.047
		0.4	0.056	0.07	0.05	0.066	0.053	0.055	0.066
	15	0.1	0.059	0.065	0.046	0.054	0.059	0.051	0.059
		0.2	0.061	0.052	0.05	0.069	0.062	0.052	0.061
		0.3	0.06	0.07	0.053	0.055	0.069	0.072	0.05
		0.4	0.079	0.077	0.049	0.064	0.049	0.055	0.066
	25	0.1	0.057	0.057	0.051	0.055	0.071	0.054	0.072
		0.2	0.064	0.07	0.06	0.064	0.051	0.064	0.056
		0.3	0.064	0.076	0.053	0.06	0.064	0.063	0.057
		0.4	0.069	0.073	0.046	0.063	0.067	0.055	0.053
	30	0.1	0.065	0.072	0.036	0.059	0.062	0.053	0.05
		0.2	0.063	0.069	0.055	0.05	0.058	0.054	0.052
		0.3	0.06	0.076	0.053	0.047	0.066	0.048	0.057
		0.4	0.058	0.054	0.053	0.062	0.073	0.052	0.063
50	5	0.1	0.058	0.073	0.049	0.067	0.059	0.045	0.066
		0.2	0.06	0.051	0.04	0.068	0.052	0.059	0.054
		0.3	0.067	0.066	0.053	0.062	0.068	0.057	0.056
		0.4	0.059	0.073	0.045	0.061	0.062	0.049	0.061
	15	0.1	0.07	0.065	0.047	0.055	0.057	0.055	0.062
		0.2	0.06	0.058	0.054	0.062	0.057	0.049	0.05
		0.3	0.06	0.062	0.058	0.07	0.061	0.049	0.049
		0.4	0.053	0.056	0.046	0.062	0.052	0.073	0.052
	25	0.1	0.056	0.059	0.05	0.057	0.053	0.063	0.052
		0.2	0.054	0.071	0.048	0.068	0.071	0.056	0.06
		0.3	0.053	0.052	0.053	0.046	0.051	0.064	0.064
		0.4	0.067	0.073	0.048	0.045	0.058	0.051	0.05
	30	0.1	0.05	0.053	0.044	0.044	0.058	0.067	0.057
		0.2	0.054	0.064	0.056	0.058	0.065	0.05	0.054
		0.3	0.065	0.061	0.051	0.052	0.058	0.06	0.054
		0.4	0.069	0.056	0.062	0.063	0.059	0.049	0.07

4 Discussion and Conclusion

In this study, we investigated the effect of sample sizes on the performance of parameter estimates at the group-level and individual-level of a two-level regression model. To check the accuracy of estimates of fixed coefficients and variance components, we considered profile likelihood method which is a valid procedure in both situations when the sample size is small and variance components are to be estimated. In our study, we have followed the simulation design presented by Maas and Hox (2005). Hence, our results are comparable with their findings. With a group of 50, they found the non-coverage rates for fixed effect parameters range from 5.1% to 5.7% and for level-2 variance components from 7.2% to 7.4%. They also mentioned that although the non-coverage rates for level-2 variance components are different from the nominal 5%, these can be considered acceptable. They reported a minimum of 50 groups to produce reliable parameter estimates.

From our simulation results, we find that the biases of estimates of the regression coefficients and variance components are not substantial. Again, when the number of groups is 30, the non-coverage rates for σ_{u0} and σ_{u1} are 6.8% and 7%, respectively. Also, the non-coverage rates for the fixed effect parameters are also reasonable for 30 groups. Consequently, from our findings, we recommend a minimum of 30 groups with 15 units per each group for drawing a valid conclusion about the population. In our study, we assumed group-level residuals are normally distributed. This may not be the case always. The consequence of violation of normality assumption of group-level residuals is studied by Maas and Hox (2004).

Declarations

Funding: None.
Conflict of interest: We declare that we have no conflict of interest.
Ethical approval: The study is based on simulated data and hence we do not need to take any ethical approval.

References

Bates DM (2010). *lme4: Mixed-effects modeling with R*. Berlin.

Berkhof J, Snijders TA (2001). "Variance component testing in multilevel models." *Journal of Educational and Behavioral Statistics*, **26**(2), 133–152. doi:10.3102/10769986026002133.

Browne WJ (1998). *Applying MCMC methods to multi-level models*. Ph.D. thesis, Citeseer.

Busing F (1993). "Distribution characteristics of variance estimates in two-level models." *Preprint PRM*, pp. 93–04.

Clarke P, Wheaton B (2007). "Addressing data sparseness in contextual population research: Using cluster analysis to create synthetic neighborhoods." *Sociological methods & research*, **35**(3), 311–351. doi:10.1177/0049124106292362.

Cohen J (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: erlbaum.

Gerhard D (2010). *Simultaneous Small Sample Inference Based on Profile Likelihood*. Ph.D. thesis.

Goldstein H (2011). *Multilevel statistical models*, volume 922. John Wiley & Sons.

Hox J (1998). "Multilevel modeling: When and why." In *Classification, data analysis, and data highways*, pp. 147–154. Springer.

Hox JJ, Moerbeek M, Van de Schoot R (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Khan MHR, Shaw E (2011). "Multilevel Logistic Regression Analysis Applied to Binary Contraceptive Prevalence Data." *Journal of Data Science*, **9**(1), 93–110. doi:10.6339/JDS.2011.09(1).647.

Kreft IG (1996). "Are multilevel techniques necessary? An overview, including simulation studies." *Unpublished manuscript, California State University, Los Angeles*.

Maas CJ, Hox JJ (2004). "Robustness issues in multilevel regression analysis." *Statistica Neerlandica*, **58**(2), 127–137. doi:10.1046/j.0039-0402.2003.00252.x.

Maas CJ, Hox JJ (2005). "Sufficient sample sizes for multilevel modeling." *Methodology*, **1**(3), 86–92. doi:10.1027/1614-2241.1.3.86.

Manor O, Zucker DM (2004). "Small sample inference for the fixed effects in the mixed linear model." *Computational statistics & data analysis*, **46**(4), 801–817. doi:10.1016/j.csda.2003.10.005.

Van der Leeden R, Busing F, Meijer E (1997). "Applications of bootstrap methods for two-level models." In *Multilevel Conference, Amsterdam*.

Welham S, Thompson R (1997). "Likelihood ratio tests for fixed model terms using residual maximum likelihood." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **59**(3), 701–714. doi:10.1111/1467-9868.00092.