Journal of Statistical Research *ISSN 0256 - 422 X*
2009, Vol. 43, No. 1, pp. 69-88
Bangladesh

COMPARISON OF RESTRICTED MAXIMUM LIKELIHOOD AND BOOTSTRAP VIA MINIMUM NORM QUADRATIC UNBIASED ESTIMATORS FOR HIERARCHICAL LINEAR MODELS UNDER χ_1^2 ASSUMPTIONS

Ayesha Delpish
Elon University, 2320 CB, Elon, NC 27244, USA
Email: adelpish@elon.edu

SUMMARY

This study investigates whether the bootstrap via minimum norm quadratic estimation procedure offers improved accuracy in the estimation of the parameters and their standard errors for a two-level hierarchical linear model when the observations follow a χ_1^2 distribution. Through Monte Carlo simulations, the importance of this assumption for the accuracy of multilevel parameter estimates and their standard errors is assessed using the accuracy index of absolute relative bias and by observing the coverage percentages of 95% confidence intervals constructed for both estimation procedures. Study results show that while both the restricted maximum likelihood and the bootstrap via MINQUE estimates of the fixed effects were accurate, the efficiencies of the estimates were affected by the distribution of errors with both procedures producing less efficient estimators under the χ_1^2 distribution, particularly for the variance-covariance component estimates.

Keywords and phrases: hierarchical linear model; multilevel model; bootstrap; minimum norm quadratic unbiased estimator

AMS Classification: 62G05; 62G09

1 Introduction

Hierarchical liner models (HLM) are often used in fields such as educational and organizational research, economics, and biology. The models are also referred to as variance or covariance components models (Dempster et al., 1981), random coefficients models (Rosenberg, 1973), multilevel linear models (Mason et al., 1983), mixed-effects and random-effects models (Laird & Ware, 1982), and mixed linear models (Goldstein, 1986). Hierarchical linear models are subsumed under the rubric of the mixed linear models (Davidian & Giltinan, 1995) and can be considered as extensions of standard linear regression models (Paterson & Goldstein, 1991). Like the ordinary multiple regression model, the underlying assumptions for a hierarchical linear model are linearity, normality of the residuals and homoscedasticity.

In hierarchical linear models, the maximum likelihood and the restricted maximum likelihood estimation procedures are often used for estimation–both of which rely on the assumptions of normality and large-sample theory. It is known that data analyzed by way of procedures that depend on normality, such as the maximum likelihood approaches, can have serious implications for the conclusions reached from (mis)using such inferential techniques. An alternative estimator to the maximum likelihood procedures is Rao's (1971) *minimum norm quadratic unbiased estimator* (MINQUE). The MINQUE is an attractive estimator mainly because the theory is developed without reference to normality. The drawback to the procedure however, is that no formulae exist for the computation of the standard errors of the MINQUE estimators (Bagakas, 1992). Thus the MINQUE can not be used directly to perform traditional statistical inference. In such situations, the bootstrap procedure is useful.

The bootstrap is a resampling approach in which the sampling properties of a statistic are examined by recomputing its value based on resamples from the original sample (Efron & Stein, 1981). The bootstrap can be carried out either parametrically or nonparametrically via estimation techniques such as maximum likelihood or MINQUE to estimate parameters and their corresponding standard errors, construct confidence intervals, and obtain the sampling distribution of the statistic. The nonparametric bootstrap is used in particular when one cannot rely on the assumption of normality. The bootstrapping of the MINQUE estimator is thus an estimation method that does not rely on the normality assumption. Throughout the paper this partnership will be referred to as the bootstrap via MINQUE procedure.

The purpose of the study is to investigate the relative performance of two estimation procedures, the restricted maximum likelihood and the bootstrap via MINQUE, for a two-level hierarchical linear model when normality conditions are not met. Specific focus lies on observing whether the bootstrap via MINQUE procedure offers improved accuracy in the estimation of the model parameters and their standard errors under the χ_1^2 distribution. The χ_1^2 distribution is a markedly skewed distribution and thus represents a large departure from normality.

2 The Two-level Hierarchical Linear Model

We begin by specifying the level-1 model. Assume that we have N subjects naturally grouped into J units. Within each unit j, there are n_j subjects with $\sum_{j=1}^{J} n_j = N$. Assume further that for the J units, the response n_j-vector is modeled with

$$\boldsymbol{Y}_j \quad = \quad \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j, \tag{2.1}$$

where $\boldsymbol{Y}_j$ is an $n_j \times 1$ vector of outcomes, $\boldsymbol{X}_j$ is an $n_j \times q$ matrix of predictor variables, $\boldsymbol{\beta}_j$ is a $q \times 1$ vector of unknown parameters and $\boldsymbol{\epsilon}_j$ is the error of prediction of $\boldsymbol{Y}_j$ by the $\boldsymbol{X}$'s. Also assume $\boldsymbol{\epsilon}_j \sim N(\boldsymbol{0}, \sigma_j^2\boldsymbol{R}_j)$ where $\boldsymbol{R}_j$ is a positive definite matrix known up to a parameter ρ_j. For simplicity, homoscedasticity is often assumed; that is, $\boldsymbol{R}_j = \boldsymbol{I}_{n_j}$ with $\boldsymbol{I}_{n_j}$ an $n_j \times n_j$ identity matrix.

The variation of each regression coefficient in $\boldsymbol{\beta}_j$ from one unit to another is modeled at level-2. Assume that information on the jth unit which is relevant to the kth element of $\boldsymbol{\beta}_j$ is denoted $\boldsymbol{\omega}_{jk}$. Then

$$\boldsymbol{\beta}_{jk} = \boldsymbol{\omega}'_{jk}\boldsymbol{\gamma}_j + \boldsymbol{\delta}_{jk}.$$

Since such an equation may be formulated for each of the q elements of $\boldsymbol{\beta}$, by writing

$$\begin{aligned}
\boldsymbol{W}_j &= \begin{pmatrix} \boldsymbol{\omega}'_{j1} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\omega}'_{j2} & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\omega}'_{jq} \end{pmatrix} \\
&= \boldsymbol{\omega}'_{j1} \oplus \boldsymbol{\omega}'_{j2} \oplus \cdots \oplus \boldsymbol{\omega}'_{jq} \\
\\
\boldsymbol{\gamma} &= (\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_2, \cdots, \boldsymbol{\gamma}'_q)' \\
\boldsymbol{\delta}_j &= (\boldsymbol{\delta}_{j1}, \boldsymbol{\delta}_{j2}, \cdots, \boldsymbol{\delta}_{jq})'
\end{aligned}$$

we obtain the level-2 model

$$\boldsymbol{\beta}_j = \boldsymbol{W}_j\boldsymbol{\gamma} + \boldsymbol{\delta_j}, \tag{2.2}$$

where $\boldsymbol{W}_j$ is a $q \times p$ matrix of predictors, $\boldsymbol{\gamma}$ is a $p \times 1$ vector of fixed effects and $\boldsymbol{\delta_j}$ is a $q \times 1$ vector of level-2 errors or random effects which is normally distributed with $E(\boldsymbol{\delta}_j) = 0$, and $Cov(\boldsymbol{\delta}_j) = \boldsymbol{\Delta}_j = \boldsymbol{\sigma}_j^2\boldsymbol{D}$. We also impose that $Cov(\boldsymbol{\delta}_j, \boldsymbol{\delta}_{j'}) = 0$, $Cov(\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_{j'}) = 0$, and $Cov(\boldsymbol{\delta}_j, \boldsymbol{\epsilon}_{j''}) = 0$, for any j, j', and j'' with $j \neq j'$. By substituting Equation (2.2) into Equation (2.1), the full multilevel regression model equation can be specified as

$$\boldsymbol{Y}_j = \boldsymbol{X}_j\boldsymbol{W}_j\boldsymbol{\gamma} + \boldsymbol{X}_j\boldsymbol{\delta}_j + \boldsymbol{\epsilon}_j. \tag{2.3}$$

2.1 Model Assumptions

Two levels of distributional assumptions can be specified for the two-level model as described in Equation (2.3). At level-1, as presented in Equation (2.1), each $\boldsymbol{\epsilon}_j$ is assumed to be independently and normally distributed with mean vector 0 and variance $\sigma^2\boldsymbol{I}_{n_j}$ for $j = 1, \ldots, J$ assumed common across groups; that is, $\boldsymbol{\epsilon}_{ij} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n_j})$. The level-1 predictors, $\boldsymbol{X}_j$, are also assumed independent of the level-1 random effects; that is, $Cov(\boldsymbol{X}_{qj}, \boldsymbol{\epsilon}_{qj}) = 0$ for all q.

At the second level of the model, as presented in Equation (2.2), it is assumed that

$$\boldsymbol{\delta}_j = (\boldsymbol{\delta}_{j1}, \boldsymbol{\delta}_{j2}, \ldots, \boldsymbol{\delta}_{jq})' \overset{iid}{\sim} N(0, \boldsymbol{T}).$$

For a model with one predictor at each level, the level-2 random errors can thus be specified by

$$\boldsymbol{\delta}_j = \begin{bmatrix} \delta_{0j} \\ \delta_{1j} \end{bmatrix} \quad \text{with} \quad \boldsymbol{T} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix},$$

where τ_{00} is the variance of the level-2 intercept δ_{0j}, τ_{11} is the variance of the level-2 slope δ_{1j}, and the covariances between the two denoted are τ_{01} and τ_{10}.

Similar to the level-1 assumption, the level-2 predictors, $\boldsymbol{W}_j$, are assumed independent of the level-2 random errors, $\boldsymbol{\delta}_j$ [i.e., $Cov(\boldsymbol{W}_j, \boldsymbol{\delta}_j) = 0$]. The general model also assumes that the level 1 errors, $\boldsymbol{\epsilon}_j$, are independent of the level 2 errors, $\boldsymbol{\delta}_j$ [i.e., $Cov(\boldsymbol{\epsilon}_j, \boldsymbol{\delta}_j) = 0$], and that no correlation exists between the predictors at one level and the random errors at another.

2.2 Robustness

In 1986, Raudenbush & Bryk wrote:

> There has been little empirical work on the consequences of violating normal distribution assumptions in HLM, but we suspect that problems are most likely to occur in estimates of the model variances, σ^2 and $\boldsymbol{T}$, and in hypothesis testing. This suggests that one should be very cautious in making substantive inferences on the basis of these statistics. (p. 14)

Their suspicions were in fact correct. Since then the effects of the violation of the normality assumption have been addressed and it is known that while the regression coefficients and their standard errors are relatively unbiased, the variance coefficients and their standard errors are not. Raudenbush & Bryk's caution concerning inference has not been heeded however, and many researchers still proceed with computations even when normality may not be guaranteed.

In classical ANOVA, non-normality can have devastating effects on the power particularly when sampling from a heavy-tailed distribution (Wilcox, 1990). The reduction in power is due to the fact that heavy-tailed distributions inflate the variance. Many robustness studies in hierarchical linear modeling of educational research data have focused on distributions such as the t distribution. The researcher found few studies investigating the effects of heavy-tailed distributions on multilevel parameter estimation and inference. Seltzer et al. (2002) used a Markov Chain Monte Carlo algorithm to investigate the robustness of the bayesian formulation of the one-way ANOVA model with random effects (the simplest of hierarchical models) to violations of normality employing t level-1 assumptions in the presence of outliers. In 1993, Seltzer similarly detailed the recalculation of the posterior distribution of fixed effects using Gibbs Sampling approach under t level-2 distributional assumptions. Seltzer reported that under heavy-tailed level-2 assumptions, estimation accuracy was less affected by outliers compared with level-2 normality assumptions.

Additionally, when sample sizes are small, the accuracy of the estimates obtained using asymptotic methods such as maximum likelihood estimation is questioned. Snijders & Bosker (1999, p. 140) offer the following observation:

> Requirements on the sample size at the highest level, for a hierarchical linear model with q explanatory variables at this level, are at least as stringent as requirements on the sample size in a single level design with q explanatory variables.

Like Kreft (1996), they suggest 30 as the optimal minimum number of groups and group size required to obtain precise estimates, particularly of the standard errors of the variance components. These findings were also supported by Basiri (1988) who concluded that it was indeed more important to have a large number of groups rather than a large number of individuals per group. His study found that not only did a larger number of groups result in increased accuracy of level-2 estimates but also that another factor, the intraclass correlation, exerted a significant effect on the accuracy of the error estimates.

The intraclass correlation coefficient (ICC), ρ, expresses the amount of relatedness of the observations within a group, such as a school (Goldstein, 1995), and is defined as the ratio of the variance component due to groups to the total variance. According to Turner in Shackman (2001, p. 2), the intraclass correlation "represents the likelihood that two elements in the same [group] have the same value, for a given statistic, relative to two elements chosen completely at random in the population ... A value of 0.05 ... is interpreted, therefore, to mean that the elements in the [group] are about 5% more likely to have the same value than if the two elements were chosen at random in the survey. The smaller the value, the better the overall reliability of the sample estimate will be". In the presence of such dependencies, Pedhazur (1997) found that the standard errors of the level-1 coefficients were often underestimated even when the intraclass correlation was very small.

Based on these cautions, the minimum sample size used for this study was 30 at each level, and the ICC was set to 0.20 since most ICCs were found to have values below 0.20 in multilevel research (Gulliford et al., 1999).

3 Estimation

In multilevel estimation, full maximum likelihood (FML) and restricted maximum likelihood (REML) estimation are the most commonly used procedures for estimation. Even though both these methods rely on classical asymptotic theory and assume normally distributed errors, their estimation results are often different.

3.1 Maximum Likelihood Estimation: FML versus REML

Maximum likelihood estimators for mixed models were first discussed by Hartley & Rao (1967) with later computational developments often based on Dempster et al.'s (1977) expectation-maximization (EM) algorithm. Other algorithms proposed for obtaining the maximum likelihood estimates include Longford's (1987) Fisher scoring methods and Goldstein's (1986) iterative generalized least-squares (IGLS) algorithms.

Maximum likelihood estimators are those estimates for which the likelihood of observing the data $\boldsymbol{Y}$ is a maximum (Raudenbush & Bryk, 2002). Consider the combined single equation model provided in Equation (2.3) which is repeated here for convenience:

$$\mathbf{Y}_j = \boldsymbol{X}_j \boldsymbol{W}_j \boldsymbol{\gamma} + \boldsymbol{X}_j \boldsymbol{\delta}_j + \boldsymbol{\epsilon}_j$$

with $\boldsymbol{\epsilon}_j \sim N(0, \sigma^2 I_{n_j})$ and $\boldsymbol{\delta}_j \sim N(0, \boldsymbol{T})$. This model may be viewed as a special case of the mixed linear model with fixed effects $\boldsymbol{\gamma}$ and random effects $\boldsymbol{\delta_j}$. $E(\boldsymbol{Y}_j) = \boldsymbol{X}_j \boldsymbol{W}_j \boldsymbol{\gamma}$ and $\boldsymbol{V}_j = \boldsymbol{X}_j \boldsymbol{T} \boldsymbol{X}_j' + \sigma^2 \boldsymbol{I}$ where $\boldsymbol{V}_j$ is the dispersion matrix and $\boldsymbol{T}$ is the variance-covariance matrix. As detailed in Afshartous & de Leeuw (2004), the full log-likelihood for the jth unit of the general two-level hierarchical linear model is

$$L_j(\sigma^2, \boldsymbol{T}, \boldsymbol{\gamma}) = -\frac{n_j}{2} log(2\pi) - \frac{1}{2} log|\boldsymbol{V}_j| - \frac{1}{2} \mathbf{d_j'} \boldsymbol{V}_\mathbf{j}^{-\mathbf{1}} \mathbf{d_j}, \tag{3.1}$$

where $\mathbf{d_j} = \boldsymbol{Y}_\mathbf{j} - \boldsymbol{X}_\mathbf{j} \boldsymbol{W}_\mathbf{j} \gamma$.

The log-likelihood for the entire model can be written as

$$L(\sigma^2, \boldsymbol{T}, \boldsymbol{\gamma}) = \sum_{j=1}^{J} L_j(\sigma^2, \boldsymbol{T}, \boldsymbol{\gamma}), \tag{3.2}$$

where the log-likelihood is expressed as a sum of unit log-likelihoods in view of the fact that the J units are independent. If the model is not approximately true and the sample size is not very large, maximizing this likelihood function for a 2-level hierarchical linear model is very difficult since in most cases the maximizer of the likelihood cannot be written in a closed form (Eliason, 1993; Raudenbush & Bryk, 2002)—hence the need for iterative schemes such as the EM-algorithm, Fisher scoring and IGLS mentioned previously.

In multilevel estimation, full maximum likelihood (FML) and restricted maximum likelihood (REML) functions are commonly used for variance component analysis. The difference in the two approaches lies primarily in the treatment of the likelihood—the variance components are estimated by the values that maximize the likelihood function over the parameter space in the full maximum likelihood procedure, while restricted maximum likelihood partitions the likelihood into pieces and maximizes the portion which is free of the fixed effects. Although the full maximum likelihood estimators of the fixed regression coefficients for a general class of regression models, including multilevel models, were proved to be unbiased (Magnus, 1978), simulation studies by Busing (1993) revealed that the variance components estimates in the hierarchical linear model obtained via this method are downward biased—this can often suggest more precision to the researcher than actually exists.

Through Monte Carlo simulations van der Leeden et al. (1997) found that although the standard errors of the variance components are generally estimated too small for both full maximum likelihood and restricted maximum likelihood, the restricted maximum likelihood estimates were more accurate. These results were confirmed by Browne (1998) who concluded that the restricted maximum likelihood estimator is almost always at least as good as the full maximum likelihood and sometimes better, especially in estimating variance components, and Harville (1977) who recommended the use of restricted maximum likelihood estimating equations when one cannot rely on the assumption of normality. Thus the restricted maximum likelihood estimation procedure was used in this study.

It is crucial to note that the restricted maximum likelihood estimation procedure still relies on the normality assumption which may not be guaranteed in most multilevel data applications—an assumption which is not required for estimation using Rao's MINQUE (1970, 1971a, 1971b).

3.2 MINQUE in HLM

Rao (1970, 1971a, 1971b) proposed the minimum norm quadratic unbiased estimator (MINQUE) whose theory was developed without reference to normality or variance of the estimators. In practice, the estimator is often used with estimated generalized least squares estimators (EGLS) to obtain estimates of the fixed model parameters (Swallow & Monahan, 1984). The MINQUE technique involves estimating a linear function of the variance components,$P'\sigma$, using a quadratic function of the observations $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}$ assumed to possess the properties of unbiasedness, translation invariance and minimum norm for the vector of observations $\boldsymbol{Y}$ and a symmetric matrix, $\boldsymbol{A}$ (Searle, 1979).

Bagakas (1992)detailed Rao's general MINQUE estimators and the derivation for hierarchical linear models as follows. Begin by recalling the combined two-level model presented in Equation (2.3). With $\boldsymbol{Z} = \boldsymbol{X}_j\boldsymbol{W}_j$, we can rewrite the equation as

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{\epsilon}, \tag{3.3}$$

where:

- $\boldsymbol{Y}$ is an $(n \times 1)$ vector of n observations;
- $\boldsymbol{Z}$ is an $(n \times P)$ known matrix of rank $r(Z) < n$;
- $\boldsymbol{\gamma}$ is a $(P \times 1)$ vector of P fixed effects parameters;
- $\boldsymbol{X}$ is an $(n \times J)$ known matrix;
- $\boldsymbol{\delta}$ is a $(J \times 1)$ vector of J unobservable random effects parameters;
- $\boldsymbol{\epsilon}$ is a $(n \times 1)$ vector of random error terms;

The vector $\boldsymbol{\delta}$ is partitioned as

$$\boldsymbol{\delta}' = [\boldsymbol{\delta}'_1 \ldots \boldsymbol{\delta}'_k \ldots \boldsymbol{\delta}'_c] \tag{3.4}$$

in order to identify the variance components for each random effect, $\boldsymbol{\delta}$. Similarly, the matrix $\boldsymbol{X}$ is partitioned as

$$\boldsymbol{X} = [\boldsymbol{X}_1 \ldots \boldsymbol{X}_k \ldots \boldsymbol{X}_c] \tag{3.5}$$

such that Equation (3.3) can be rewritten in a form similar to Rao's (1971a) equation 5.3 as

$$\boldsymbol{Y} = \boldsymbol{Z\gamma} + \sum_{k=0}^{c} \boldsymbol{X}_k \boldsymbol{\delta}_k, \tag{3.6}$$

where $\boldsymbol{\epsilon}$ is written as $\boldsymbol{\delta}_0$ with corresponding $\boldsymbol{X}_0 = \boldsymbol{I}_n$.

Using this formulation, Bagakas (1992) derived the multilevel form of the vector of MINQUE estimators of the variance and covariance components as:

$$\hat{\boldsymbol{\sigma}} = \{tr(\boldsymbol{P}_w \boldsymbol{Z}_k \boldsymbol{Z}_k' \boldsymbol{P}_w \boldsymbol{Z}_{k'} \boldsymbol{Z}_{k'}')\}^{-1} \{\boldsymbol{Y}' \boldsymbol{P}_w \boldsymbol{Z}_k \boldsymbol{Z}_k' \boldsymbol{P}_w \boldsymbol{Y}\} \tag{3.7}$$

with the projector operator, $\boldsymbol{P}_w$, on the space generated by the columns of Z defined similarly to Rao's (1971b) equation 1.2 as

$$\boldsymbol{P}_w = \boldsymbol{V}_w^{-1} - \boldsymbol{V}_w^{-1} \boldsymbol{Z} (\boldsymbol{Z}' \boldsymbol{V}_w^{-1} \boldsymbol{Z})^{-1} \boldsymbol{Z}' \boldsymbol{V}_w^{-1}, \tag{3.8}$$

where $\boldsymbol{V}_w = \boldsymbol{X D}_w \boldsymbol{X}'$ and $\boldsymbol{D}_w = \text{diag}\{\boldsymbol{w}_0 \boldsymbol{I}_n, \boldsymbol{w}_1 \boldsymbol{I}_j\}$ is a matrix of δ weights w_0 and w_1 in the norm given by $w_0 = 1 - \rho$ and $w_1 = \rho$ for the intraclass correlation, ρ.

After mathematical manipulations, Bagakas (1992) also showed that the fixed effects parameters of the model can be estimated using

$$\hat{\gamma} = \boldsymbol{K Z}' \boldsymbol{V}_w^{-1} \boldsymbol{Y}, \tag{3.9}$$

which eventually is shown to have simplified form

$$\hat{\gamma} \quad = \quad w \sum (\boldsymbol{K Z}_j' \boldsymbol{Y}_j - c_j r_j \boldsymbol{K S}_j), \tag{3.10}$$

where $w = 1/(1 - w_1)$ and $c_j = w_1/(1 + (n_j - 1)w_1)$ for $j = 1, 2, \ldots, J$; $\boldsymbol{Z}_j$ is the n_j rows of the matrix $\boldsymbol{Z}$ associated with fixed effects in the jth group; $r_j = \boldsymbol{X}_{1j}' \boldsymbol{Y}_j$ is the sum of $\boldsymbol{Y}$ elements in group j; $\boldsymbol{S}_j = \boldsymbol{Z}_j' \boldsymbol{X}_{ij}$ is a $(P \times 1)$ vector of column sums of $\boldsymbol{Z}_j$ and $\boldsymbol{K}$ defined as

$$\boldsymbol{K} = w \sum (\boldsymbol{Z}_j' \boldsymbol{Z}_j - c_j \boldsymbol{S}_j \boldsymbol{S}_j'). \tag{3.11}$$

For complete details on the derivation of the above formulae, the reader is referred to Bagakas (1992) and Rao (1970, 1971a, 1971b).

Although the MINQUE procedure has been mentioned as a potential alternative to the maximum likelihood procedure (Rao & Kleffe, 1989; Swallow & Searle, 1978; Kreft et al., 1990), after an exhaustive search of the literature, few studies were found that investigate the estimator's use in multilevel modeling. In an unpublished dissertation, Bagakas (1992) compared the usual MINQUE estimates to bootstrapped estimates for a two-level random intercepts model. Using a Monte Carlo simulation study, he observed the performance of the two procedures under various conditions—normal versus Laplace errors; unbalanced group sizes and the intraclass correlation observed at three levels: 0.01, 0.05 and 0.20. He found that the bootstrap via MINQUE technique produced lower standard errors with minimal bias and allowed for the generation of bootstrap confidence intervals. Based largely on Bagakas's (1992) findings and on the bootstrap's unreliance on the distributional assumptions, the bootstrap via MINQUE procedure was adopted for this study.

4 Method

A series of Monte Carlo simulations were conducted on a two-level hierarchical linear model with one explanatory variable at each level to examine the performance of the parameter estimates and their standard errors obtained from two estimation procedures—restricted maximum likelihood and bootstrap via MINQUE. SAS and SAS/IML programs (2003) were used to generate the data, implement estimation techniques, fit the specified model and compute estimation accuracy indices for both approaches.

4.1 The Simulation Model

The two-level random coefficients model represented a reasonable model that, in the interest of parsimony, relies on a small number of variables. Formally the model is as follows:

$$Level-1: Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij} \tag{4.1}$$

$$\begin{aligned} Level-2: \beta_{0j} &= \gamma_{00} + \gamma_{01}W_j + \delta_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}W_j + \delta_{1j}, \end{aligned} \tag{4.2}$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$, and a covariance matrix $\boldsymbol{T}$ consists of τ_{00}, the variance of the level-2 intercept δ_{0j}; τ_{11}, the variance of the level-2 slope δ_{1j}; and the covariance between the two denoted τ_{01} and τ_{10}.

An observed value Y_{ij} was generated through the combined model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X_{ij} + \gamma_{11}W_jX_{ij} + (\delta_{1j}X_{ij} + \delta_{0j} + \epsilon_{ij}). \tag{4.3}$$

The following parameters are estimated based on Equation (4.3):

- Fixed-effect coefficients: $\gamma_{00}, \gamma_{10}, \gamma_{01}, \gamma_{11}$
- The level-1 variance component: σ^2
- The level-2 variance-covariance components: $\tau_{00}, \tau_{11}, \tau_{01}$

Following Cohen (1998), the intercept coefficient was set to 1.00, and a medium sized effect of 0.30 was used for the slopes. The level-1 variance component, σ^2, was set to 0.5 while the level-2 variance-covariance component, τ_{00} and τ_{01}, followed from the specification of σ^2 and the intraclass correlation coefficient, ρ. For $\rho = 0.20$, $\tau_{00} = 0.125$ and $\tau_{01} = 0.0625$. Finally, the value for the slope variance, τ_{11}, was equated to that of the intercept variance, τ_{00}, since a 1993 study by Busing showed that the effects for these coefficients were similar.

Number of Groups

Various simulation studies performed (c.f., Mok, 1995; Maas & Hox, 2005) suggest that the number of groups (J) at level-2 affects the estimation accuracy more than the group sizes (n_j) at level-1. Maas & Hox (2005) also found that having balanced or unbalanced groups had no influence on the maximum likelihood estimates and their standard errors. Attention is thus restricted to the balanced case $n_j = n$ with a value of 30 chosen based on simulation results from researchers such as Kreft (1996) and Kreft & de Leeuw (1998) who proposed a 30–30 rule—30 as the minimum required sample size at both level-1 and level-2 to ensure accurate parameter estimation.

There were two levels used for the number of groups factor (NG) with the choice based on the literature and practice. The lowest level, 30, is the smallest number of allowable groups proposed by Kreft's (1996) 30–30 rule, while 100 is sufficiently large based on results by researchers such as Swaminathan (2001) and van de Leeden et al. (1997). Thus while the level-1 or group size was fixed at 30, the level-2 size (number of groups) was varied to observe the performance of the estimators.

Intraclass Correlation

The intraclass correlation coefficient (ICC), ρ, represents the correlation between pairs of values within the J groups and measures the degree of dependence among observations (Raudenbush & Bryk, 2002). The function is given by the ratio of the level-2 variance (the variance component due to groups) to the total variance in the null representation of the model. By setting all the coefficients of the explanatory variables of Equation (4.3) to zero, we obtain the null (one-way ANOVA with random effects) representation of the model:

$$Y_{ij} = \gamma_{00} + \delta_{0j} + \epsilon_{ij}.$$

The intraclass correlation based on the one-way ANOVA with random effects model is specified as

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}, \tag{4.4}$$

where σ^2 is the variance of the level-1 residuals and τ_{00} is the variance of the level-2 intercept errors. Gulliford et al. (1999) found that most ICCs were found to have values below 0.20 in multilevel research, so this value was used for the present study. With specified $\sigma^2 = 0.5$ and $\rho = 0.20$, we obtain $\tau_{00} = \tau_{11} = 0.125$,and $\tau_{01} = 0.0625$.

The preceding simulation decisions are summarized in Table (1) below.

Parametric Assumption

A pilot study performed on a double exponential (Laplace) distribution revealed little to no significant difference between the estimates obtained from the restricted maximum likelihood procedure and those from the bootstrap via MINQUE procedure. The double exponential distribution represented the case where the distribution was still symmetric although the tails are longer and thinner than those of the normal distribution. It appeared to the researcher that the Laplace distribution did not differ enough from the normal distribution to unambiguously show the benefits of the bootstrap procedures. The χ^2_1 distribution was then adopted since it represents a markedly skewed departure from normality.

Table 1: Simulation Conditions & Population Parameters

Sample Size	
Level-2 Sample Size, J	$30, 100$
Level-1 Sample Size, n_j	30
Fixed Components	
Intercept, γ_{00}	1.00
Slopes, $\gamma_{01}, \gamma_{10}, \gamma_{11}$	0.30
Random Components	
Level-1 variance, σ^2	0.50
Level-2 variances, $\tau_{00} = \tau_{11}$	0.125
Covariance, τ_{01}	0.0625
Intraclass correlation, ρ	0.20
Distributions	χ_1^2

4.1.1 Procedure

Following Kendall & Stout, in Busing (1993), who reported that between 100 and 500 Monte Carlo replicates are adequate to be considered large enough to ensure stability of results, 500 replicates were used in each condition of this study. This resulted in 1000 trials being performed across the two condition models (χ_1^2 with NG= 30 versus χ_1^2 with NG= 100). The analyses were carried out twice—once with the restricted maximum likelihood estimation procedure and once using the bootstrap via MINQUE procedure.

Computer segments were coded to generate χ_1^2 variates and the χ_1^2 residuals were generated in part using the following steps:

1. Draw (a_{j0}, a_{j1}) from a bivariate normal distribution.

2. Set $\delta_{0j} = (a_{j0}^2 - 1)/k_i$ and $\delta_{1j} = (a_{j1}^2 - 1)/k_i$ where k_i are chosen so as to produce the τ values mentioned above.

4.2 Implementing the Bootstrap

The nonparametric cases bootstrap involves resampling from the original sample and assumes only that the original sample is a random sample from the parent population. For the Monte Carlo simulation study performed for this paper, the bootstrap was implemented in the following steps:

1. Construct the empirical distribution $\hat{F}_J$ by assigning mass $\frac{1}{J}$ to each of the level-2 units in the original sample.

2. Randomly sample with replacement from the J level-2 units.

3. For each of the selected level-2 units containing n_j level-1 units, construct distributions F_{n_j} by assigning mass $\frac{1}{n_j}$ to the j^{th} level-2 unit, for $j = 1, 2, \ldots, J$.

4. For each selected level-2 unit whose distribution was constructed in step 3 above, draw n_j level-1 units with replacement. This produced the bootstrap data set with a vector of observations denoted by $\boldsymbol{Y}^*$.

5. Determine the MINQUE estimate of the parameters of the model using the bootstrap replicated sample.

6. Repeat a large number B times to obtain a sequence of MINQUE estimates of the parameters of the model for $b = 1, 2, \ldots, B$.

7. Observe the distribution of the bootstrap parameter values. The empirical expectation of the bootstrap estimate of any given parameter in the model, say $\hat{\theta}^*_{(.)}$, is then found by averaging the $\hat{\theta}^*_b$ obtained bootstrap replicated estimates over all replications. That is,

$$\hat{\theta}^*_{(.)} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*_b. \tag{4.5}$$

   Additionally, the standard error of the bootstrap estimate of the parameter, denoted $s.e.(\hat{\theta}^*)$, can be computed as

$$s.e.(\hat{\theta}^*) = \left[\frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^*_b - \hat{\theta}^*_{(.)})^2 \right]^{\frac{1}{2}}. \tag{4.6}$$

For the bootstrap procedure, estimation is performed at each of b replications for $b = 1, \ldots, B$, large. A pilot study performed revealed that there was little improvement in the confidence intervals past $B = 1000$ so this number of replications was used in the present study.

Thus while one restricted maximum likelihood estimate (REML) will be obtained at each of the 1000 trials in the Monte Carlo simulation study, the bootstrap estimate at each trial of any given parameter in the model specified in Equation (4.3) will be the average of 1000 bootstrap replicated estimates.

4.3 Estimation Accuracy Indices and Analysis

Parameter estimates and their corresponding standard errors were obtained for each fixed-effect parameter $(\gamma_{00}, \gamma_{10}, \gamma_{01}, \gamma_{11})$ and variance-covariance component estimate $(\sigma^2, \tau_{00}, \tau_{01}, \tau_{11})$.

The index of absolute relative bias was used in this study to compare the degree of bias in the estimation of parameter values. For the REML-estimates, absolute relative bias was computed as the absolute proportion of the theoretical value from which the estimate departs from the theoretical value. That is,

$$\text{Absolute Relative Bias} = \left| \frac{(\hat{\theta} - \theta)}{\theta} \right|,$$

where $\hat{\theta}$ is the estimate of θ.

The bootstrap estimate of relative bias has the form

$$\begin{aligned} \widehat{BIAS} &= \frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}_b^* - \hat{\theta}) \\ &= \hat{\theta}_{(.)}^* - \hat{\theta}. \end{aligned} \tag{4.7}$$

The bootstrap estimate of bias allows for the construction of a bias-corrected estimator of the parameter of interest, $\hat{\theta}_{Boot}$, which in most cases simply corrects for scale (Schenker, 1985). The absolute bias-corrected form of $\hat{\theta}$ is given by

$$\begin{aligned} \hat{\theta}_{Boot} &= |\hat{\theta} - \widehat{BIAS}| \\ &= |2\hat{\theta} - \hat{\theta}_{(.)}^*|. \end{aligned} \tag{4.8}$$

Chen & Hall (2003) showed that the estimator, $\hat{\theta}_{Boot}$ given in Equation (4.8) above corrects for biases of order n^{-1} thus satisfying $E(\hat{\theta}_{Boot}) = O(n^{-2})$. The use of bias-corrected point estimates of the parameter θ, referred to as *pivoting* in the literature, has been advocated by researchers such as Hall (1990) and for use in the construction of bootstrap confidence intervals. Following Shieh & Fouladi (2003), the bias criterion were defined as follows: small to negligible when relative bias was less than 5%, moderate or medium bias when relative bias was between 5% to 20% and large bias when relative bias was greater than 20%.

To investigate the accuracy of the standard errors, 95% confidence intervals were also computed for each parameter in each replication for both estimation procedures. The percentage of replications in which the interval contained the true parameter was recorded and analyzed to investigate the degree to which the observed and nominal coverage percentages agreed. Confidence intervals are preferred to standard deviations as an indication of precision particularly when the underlying distribution is skewed or strongly non-normal, as is the case with the χ_1^2 distribution. In such cases, the standard error is not a good indication of the precision of the point estimate. Confidence intervals for the estimates obtained via restricted maximum likelihood were based on the usual large-sample approach, while bootstrap confidence intervals were constructed using the percentile method.

5 Results

The restricted maximum likelihood estimation procedure relies on the assumption that the observations are normally and independently distributed. However, as previously mentioned, the normality of the population distribution can not be guaranteed especially when modeling situations such as with school effects data. Unfortunately, in application many researchers proceed with computations under the normality assumption regardless of whether or not they are met—even when there is little doubt of the data's non-normality. The aim of this study is not to yet again outline methods for assessing non-normality, but rather to investigate the behavior of the restricted maximum likelihood estimators and the bootstrap via MINQUE procedure in a case where the normality assumption is known to not hold. The χ_1^2 distribution is a markedly skewed distribution and thus represents a large departure from normality.

5.1 Parameter Estimates and Relative Bias

Table(2) shows the mean absolute relative biases of the fixed and random parameter estimates. Overall the fixed effects parameter estimates were almost consistently unbiased with the REML estimates competing with those obtained via the bootstrap procedure, on average. Mean relative biases for all fixed effect estimates were negligible (magnitude <5%) with the largest fixed effect mean relative bias magnitude observed at 0.5% for the γ_{10} parameter under REML, while the largest bias under the bootstrap procedure was 0.1%. The largest biases were noted in conditions with the low level of the number of groups factor.

Table 2: Mean absolute relative bias of the fixed & random parameter estimates

Fixed		γ_{00}	γ_{10}	γ_{01}	γ_{11}
NG=30	REML	0.003	0.005	0.002	0.004
	BOOT	0.001	0.001	0.000	0.001
NG=100	REML	0.001	0.001	0.001	0.001
	BOOT	0.000	0.000	0.000	0.000
Var-Covariance		τ_{00}	τ_{11}	τ_{01}	σ^2
NG=30	REML	0.047	0.048	0.045	0.003
	BOOT	0.007	0.009	0.006	0.001
NG=100	REML	0.014	0.019	0.010	0.001
	BOOT	0.004	0.004	0.003	0.000

Although for both the REML and bootstrap procedures as number of groups increased the bias decreased in magnitude, this increase had no statistically significant effect on the absolute relative bias of the fixed parameter estimates (p-values < 0.001).When $NG = 30$, the difference in the mean absolute relative bias of the estimates obtained from the two methods is greater than when the number of groups is increased to 100. Wilcoxon signed rank tests of the pairwise differences in the mean absolute relative biases for the two methods revealed no statistically significant differences between the performance of the REML and the bootstrap procedure in the estimation of the fixed effect parameters for the χ_1^2 data at either level of the group condition (p-values $= 0.1250$).

Like the fixed effects, both estimation procedures produced negligible biases on average for all variance-covariance component parameter estimates. The largest random effects bias under REML had magnitude of approximately 5% corresponding to the τ_{11} parameter estimate when $NG = 30$, while the largest mean relative bias of the variance-covariance components observed for the bootstrap procedure was 0.9% corresponding to τ_{11} when $NG = 30$. As observed with the fixed effects, the magnitude of the mean absolute relative bias of the variance-covariance component estimates decreased as the number of groups increased for both procedures. The decrease in the absolute relative bias was larger on average for the level-2 variance-covariance parameter estimates than for the level-1 variance σ^2 for both procedures.

A comparison of the estimation procedures was then conducted using a Wilcoxon signed rank test of the pairs across conditions and revealed that there was a statistically significant difference between the estimates of the variance-covariance components obtained via the two estimation procedures with an observed p-values of 0.0078. However results suggest that both estimation procedures performed satisfactorily for the estimation of the fixed effects and the variance-covariance components.

Results thus suggest that both estimation procedures performed satisfactorily for the estimation of the fixed effects and the variance-covariance components.

5.2 Standard Errors and Coverage Probabilities

To investigate the accuracy of the standard errors under the χ_1^2 distribution, 95% confidence intervals were constructed for each parameter in each replication for both estimation procedures. Confidence intervals are preferred to standard errors as an indication of precision particularly when the underlying distribution is skewed or strongly non-normal, as is the case with the χ_1^2 distribution, so the bootstrap confidence intervals were expected to excel where the REML-intervals usually fail. The coverage rates of the 95% confidence intervals are given in Table (3) and suggest that overall the bootstrap intervals achieved better coverage rates, particularly for the variance-covariance components.

For the fixed effects parameters, the bootstrap produced coverage rates were higher than those obtained from the REML technique over simulated conditions (p-value $= 0.0048$). The lowest coverage rate observed under REML was approximately 91% for γ_{10} when $NG = 30$ compared to the lowest bootstrap rate of about 93%. For the variance-covariance parameters, as expected, the confidence intervals based on REML estimates provided insufficient

Table 3: Coverage of the 95% confidence interval for the fixed & random effects

Fixed		γ_{00}	γ_{10}	γ_{01}	γ_{11}
NG=30	REML	0.920	0.909	0.937	0.951
	BOOT	0.943	0.944	0.947	0.944
NG=100	REML	0.940	0.934	0.946	0.948
	BOOT	0.948	0.949	0.949	0.950
Var-Covariance		τ_{00}	τ_{11}	τ_{01}	σ^2
NG=30	REML	0.600	0.611	0.687	0.943
	BOOT	0.944	0.945	0.945	0.948
NG=100	REML	0.682	0.652	0.725	0.948
	BOOT	0.948	0.949	0.949	0.950

coverage for the estimates of the level-2 parameters with coverage rates ranging from approximately 60% for τ_{00} when $NG = 30$, to approximately 73% for τ_{01} when $NG = 100$. In contrast, the coverage percentages obtained by the bootstrap generated confidence intervals for the level-2 parameter estimates performed well with coverage rates ranging from 94.4% for τ_{00} when $NG = 30$ and to 94.9% for τ_{11} and τ_{01} when $NG = 100$. The level-1 variance, σ^2 performed similarly for both methods across all conditions with adequate coverage rates attained.

6 Summary and Discussion

Based on the analysis of the mean relative biases, the skewness of the underlying distribution had little to no effect on the estimates of the fixed effects obtained via the restricted maximum likelihood estimation procedure, with negligible($< 5\%$) relative biases obtained. The bootstrap procedure also produced unbiased estimates of the fixed parameters. On average, the fixed effects parameter estimates based on the bootstrap procedure performed better than those obtained via restricted maximum likelihood estimation. The coverage percentages for the fixed parameters based on the usual large-sample approach of the REML-intervals did not perform as well as the bootstrap confidence intervals.

For the variance-covariance components, the estimates from both procedures produced negligible biases. Across all conditions, estimated confidence intervals for the level-2 variance-covariance parameters produced via REML-estimation were unacceptable. Thus when normality is violated (particularly when the data are highly skewed), the REML-estimates of the standard errors should not be used for inference. These findings support Raudenbush & Bryk's (1986, p. 14) conjecture that:

> on the consequences of violating normal distribution assumptions in HLM, ... problems are most likely to occur in estimates of the model variances, σ^2 and $\boldsymbol{T}$, and in hypothesis testing. This suggests that one should be very cautious in making substantive inferences on the basis of these statistics.

Based on the results of this study, the bootstrap via MINQUE appears to be an attractive alternative to estimation in cases where normality is not guaranteed. Maas & Hox (2004, p. 439) suggest that if the normality assumption is violated, "a different approach that merits the analysts' attention is the non-parametric bootstrap". It is hoped that the nonparametric cases bootstrap via MINQUE presented in this paper will offer another useful alternative to maximum likelihood estimation procedures in hierarchical linear modeling particularly when the underlying distribution is non-normal or the sample size at level-2 is small.

References

[50] Afshartous, D., and de Leeuw, J. (2004). An application of multilevel model prediction to NELS:88. *Behaviormetrika,* **31***(1), 43–66.*

[50] Bagakas, J. G. (1992). *Two level nested hierarchical linear model with random intercepts via the bootstrap.* Unpublished doctoral dissertation, Michigan State University.

[50] Basiri, D. (1988). *Large and small sample properties of maximum likelihood estimates for the hierarchical linear model.* Ph.D. thesis, Department of Counselling, Educational Psychology and Special Education, Michigan State University.

[50] Browne, W. J. (1998). *Applying MCMC methods to multilevel models.* Unpublished doctoral dissertation, University of Bath, UK.

[50] Busing, F. M. T. A. (1993). *Distribution characteristics of variance estimates in two-level models.* Preprint PRM 93-04, Psychometrics and Research Methodology, Leiden, Netherlands.

[50] Chen, S. X., and Hall, R. (2003). Effects of bagging and bias correction on estimators defined by estimating equations. *Statistica Sinica,* **13***, 97–109.*

[50] Cohen, M. P. (1998). Determining sample sizes for surveys with data analyzed by hierarchical linear models. *Journal of Official Statistics,* **14***(3), 267–275.*

[50] Davidian, M., and Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data.* New York: Chapman & Hall/CRC.

[50] Dempster, A. P., Laird, N. M, and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B,* **39***, 1–8.*

[50] Dempster, A. P., Rubin, D. B., and Tsutakawa, R. D. (1981). Estimation in covariance components models. *Journal of the American Statistical Association,* **76***, 341–353.*

[50] Efron, B., and Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics,* **9***, 586–596.*

[50] Eliason, S. R. (1993). *Maximum Likelihood Estimation.* Newbury, CA: Sage.

[50] Goldstein, H. (1986). Multilevel mixed linear analysis using iterative generalized least squares. *Biometrika,* **73***, 43–56.*

[50] Goldstein, H. (1995). *Multilevel Statistical Models.* New York, NY: Halstead Press.

[50] Gulliford, M. C., Ukoumunne, O. C., and Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies. *American Journal of Epidemiology,* **149***, 876–883.*

[50] Hall, P. (1990). Asymptotic properties of the bootstrap for Heavy-Tailed Distributions. *The Annals of Probability,* **18***, 1342–1360.*

[50] Hartley, H. O., and Rao, J. N. K. (1967). Maximum likelihood analysis for the mixed analysis of variance model. *Biometrika,* **54***, 93–108.*

[50] Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association,* **72***, 320–340.*

[50] Kreft, I. G. G. (1996). *Are multilevel techniques necessary? An overview including simulation studies.* Unpublished report. Los Angeles, CA: University of California, Department of Statistics.

[50] Kreft, I. G. G., and de Leeuw, J. (1998). *Introducing Multilevel Modeling.* Thousand Oaks, CA: Sage.

[50] Kreft, I. G. G., de Leeuw, J., and Kim, K-S. (1990). *Comparing four different statistical packages for hierarchical linear regression: GENMOD, HLM, ML2, VARCL.* Preprint 50, UCLA Statistics, Los Angeles, CA.

[50] Laird, N. M., and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics,* **38***, 963–974.*

[50] Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. *Biometrika,* **74***, 817–827.*

[50] Maas, C. J. M., and Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics and Data Analysis,* **46***(3), 427–440.*

[50] Maas, C. J. M., and Hox, J. J. (2005). Sufficient Sample sizes for Multilevel Modeling. *Methodology,* **1***(3), 85–91.*

[50] Magnus, J. R. (1978). Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics,* **7***, 281–312.*

[50] Mason, W. M., Wong, G. Y, and Entwistle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 72–103). San Francisco: Jossey-Bass.

[50] Mok, M. (1995). Sample size requirements for 2-level designs in educational research. *Multilevel Modelling Newsletter,* **7***(2), 11–15.*

[50] Patterson, L., and Goldstein, H. (1991). New statistical methods for analyzing social structures: An introduction to multilevel models. *British Educational Research Journal,* **17***, 387–393.*

[50] Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research* (3^{rd} ed.). Orlando, FL: Harcourt Brace.

[50] Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association,* **65***, 161–172.*

[50] Rao, C. R. (1971a). Estimation of variance and covariance components–MINQUE theory. *Journal of Multivariate Analysis,* **1***, 257–275.*

[50] Rao, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis,* **1***, 445–456.*

[50] Rao, C. R., and Kleffe, J. (1989). *Estimation of variance components.* North Holland: Amsterdam.

[50] Raudenbush, S. W., and Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education,* **59***(1), 1–17.*

[50] Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2^{nd} ed.). Thousand Oaks, CA: Sage Publications.

[50] Rosenberg, B. (1973). Linear regression with randomly dispersed parameters. *Biometrika, bf 60, 61–75.*

[50] SAS Institute Inc. (2003). *SAS/STAT User's Guide, Release 9.1 Edition.* Cary, NC: SAS Institute Inc.

[50] Schenker, N. (1985). Qualms about bootstrap confidence intervals. *Journal of the American Statistical Association,* **80***, 360–361.*

[50] Searle, S. R. (1979). *Note on variance component estimation: A detailed account of maximum likelihood and kindred methodology.* Paper No. BU-673-M, Biometric Unit, Cornell University.

[50] Seltzer, M. (1993). Sensitivity analysis for fixed effects in the hierarchical model: A Gibbs sampling approach. *Journal of Educational Statistics,* **18***, 207–235.*

[50] Seltzer, M., Novak, J., Choi, K., and Linn, N. (2002). Sensitivity analysis for hierarchical models employing t level-1 assumptions. *Journal of Educational and Behavioral Statistics,* **27***(2), 181–222.*

[50] Shackman, G. (2001). *Developing Sample Sizes for New Surveys: Estimating the Design Effect Adjustment.* Paper presented at the Albany Chapter of the American Statistical Association and Rensselaer Polytechnic Institutes Applications in Statistics: What's Hot / What's Next VI. Albany, NY.

[50] Shieh, Y-Y., and Fouladi, R. T. (2003). The effect of multi-collinearity on multilevel modeling parameter estimates and standard errors. *Educational and Psychological Measurement,* **63***(6), 951–985.*

[50] Snijders, T. A. and Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage Publications.

[50] Swallow, W. H., and Monahan, J. F. (1984). Monte Carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components. *Technometrics,* **26***(1), 47–57.*

[50] Swallow, W. H., and Searle, S. R. (1978). Minimum variance quadratic unbiased estimation (MIVQUE) of variance components. *Technometrics,* **20***, 265–272.*

[50] Swaminathan, H. (2001). *Identification of factors that contribute to differential item performance among NAEP test taking populations.* US Department of Education National Center for Education Statistics, National Assessment of Educational Progress (Federal ID# 70016). University of Massachusetts: Amherst.

[50] van der Leeden, R., Busing, F., and Meijer, E. (1997). *Applications of bootstrap methods for two-level models.* Paper presented at the Multilevel Conference, Amsterdam, April 1–2.

[50] Wilcox, R. R. (1990). Comparing the means of two independent groups. *Biometrical Journal,* **32***, 771–780.*