Some Estimation Methods and Their Assessment in Multilevel Models: A Review

Yahia S El-Horbaty* and Eman M Hanafy

Department of Mathematics, Insurance and Applied Statistics, Helwan University, Egypt

Submission: January 27, 2018; **Published:** February 27, 2018

*Corresponding author: Yahia S El-Horbaty, Department of Mathematics, Insurance, and Applied Statistics, Helwan University, Egypt;

Email: yahia_mohamed@commerce.helwan.edu.eg

Abstract

Multilevel linear regression models represent a generalization of linear models in which the regression coefficients are themselves given a model whose parameters are also estimated from the data. This paper reviews multilevel random coefficients regression models with a focus on the estimation problem and its assessment. Parameter estimation for the fixed effects and the variance components are highlighted. In addition, comparisons that are made in the literature to choose among the competing methods are highlighted. This is particularly emphasized when some of the assumptions underlying the estimation methods are violated.

Keywords: Multilevel models; Parametric estimation; Robust standard errors

Abbreviations: GLS: Generalized Least Squares; MINQUE: Minimum Norm Quadratic Estimation; ML: Maximum Likelihood; REML: Restricted Maximum Likelihood; IGLS: Iterative Generalized Least Squares; EM: Expectation Maximization; OLS: Ordinary Least Squares; ARE: Asymptotic Relative Efficiency

Introduction

Multilevel modeling is an approach that can be used to handle clustered or grouped data. Social sciences often involve problems that investigate the relationship between individual and society. The general concept is that individuals and the social groups are conceptualized as a hierarchical system, where the individuals and groups are defined at separate levels of this hierarchical system. There are many types of multilevel models, which differ in terms of the number of levels, type of design (random intercept, random slopes and random coefficients regression model), scale of the outcome variable (continuous, categorical), and number of outcomes (univariate, multivariate). In this article, the two-level random coefficients regression model is represented, letting the univariate outcome variable to be continuous.

The parameters to be estimated in multilevel regression models are the fixed coefficients that represent the fixed part of the model. The parameters also include the variance covariance matrix of the random coefficients and the variances of the residual errors that represent the random part of the model. If the components of the random part in the model were known, the unknown fixed coefficients could be estimated using generalized least squares (GLS) estimation method [1]. Similarly, if the fixed coefficients were known, then the unknown

variance components of the model could also be estimated using GLS estimation method. If both are unknown, hence we have to follow different approaches.

In order to estimate the unknown parameters several procedures are discussed in Searle et al. [2]. These methods include the ANOVA method for balanced data which uses the expected mean squares approach. However, this method is difficult to deal with under unbalanced data situations. Under the latter case, Rao [3] proposed the minimum norm quadratic estimation (MINQUE) method for estimating the variance parameters that produces quadratic unbiased estimators with minimum norm (MINQUE). Maximum likelihood method (full maximum likelihood (ML) and restricted maximum likelihood (REML)), iterative generalized least squares (IGLS) and the expectation maximization (EM) estimation method represent standard methods that are used under both balanced and unbalanced data. These methods involve iterative procedures and thus the resulting estimators are not necessarily expressed in a closed form.

The commonly used methods to estimate the multilevel model are ML and REML [4]. The ML estimators of the variance components do not correct for the degrees of freedom lost due to the estimation of the fixed effects. As a result, the estimates

of the variance components are generally too small. REML estimation corrects for the uncertainty about the fixed effects in estimating the variance components, which is especially useful if the number of the second-level units (e.g. groups) is relatively small [4,5]. There are other methods used to estimate the unknown parameters of the multilevel model. An iterative two-stage procedure is the IGLS [6]. IGLS usually starts with an ordinary least squares (OLS) estimate of the fixed parameters, which is then used to estimate the random part of the model. Next, the estimate of the covariance matrix of the model is used to make an improved estimate of the fixed part, which in turn is used to improve the estimate of that covariance matrix. The IGLS method alternates between estimating the fixed and the variance components of the model until convergence is reached. As far as the maximum likelihood approach is concerned, the normality of the error distributions at each level is introduced. Goldstein [7] proved that under normality of all the error terms the parameter estimates resulting from IGLS procedure are maximum likelihood estimates.

Another method of estimation is based on the EM algorithm that developed in Dempster et al. and is applied in Raudenbush & Bryk to models covering multilevel random coefficients regression models. The EM algorithm addresses the problem of maximizing the likelihood by conceiving it as a problem in missing data where, similar to the IGLS method, its procedure is based on an iterative procedure. The aforementioned parametric estimation methods provide approximately the same results for the parameter estimates and their standard errors under standard assumptions and large sample size Goldstein [7], Raudenbush & Bryk and Gumedze & Dunne. One of the essential assumptions of the tests of significance used in multilevel programs is normality of the error distributions involved. Mass & Hox [4] highlight the fact that the aforementioned estimation methods are asymptotically efficient. The authors suggest a number of groups of 50 groups to achieve accurate estimates of the standard errors of the fixed part estimates. This suggestion is based on an earlier conclusion by simulations provided by Van der Leeden and Busing and Van der Leeden et al. in addition they suggest that when assumptions of normality and large samples are not met, the standard errors have a small downward bias. On the other hand, in dealing with outliers Longford and Lewis offered practical procedures for dealing with outliers in multilevel data then he assesses the effect of them on the model estimators using different parametric and robust parametric estimation methods.

Importantly, different measures of assessment are presented that can be used to investigate the accuracy of the parameter estimates, variances and their standard errors in multilevel regression models and compare different estimation methods. Those include the percentage relative bias, the coverage probability and the asymptotic relative efficiency. Accordingly, the rest of this article is organized as follows. Section 2 describes the two-level model. Some basic parametric and robust

parametric estimation methods are reviewed in Section 3. The measures of assessment are highlighted in Section 4 and the conclusions are set in Section 5.

Model and notations

Assume we have data from m groups, with a different number of responses n_j (where j=1,2,,...,m) in each group. Let the data be obtained in the form of a vector of the outcome variable Y_j, a set of explanatory variables in W_j, and another set of group-level explanatory variables in Z_j. To model these data, separate regression models for each level are accommodated. First, the individual-level model

$$Y_j = W_j \alpha_j + \varepsilon_j \qquad (1)$$

Where, Y_j is a vector of length n_j representing the response variable of group j, W_j is a design matrix of size $n_j \times q$, and ε_j is a vector of length n_j representing the residuals on the same level where, $\varepsilon_j \sim N\left(0, \sigma_e^2 I_{n_j}\right)$. The group-level model is given by

$$\alpha_j = Z_j \beta + \delta_j \qquad (2)$$

Where, Z_j is $q \times p$ design matrix, β is an $p \times 1$ vector of fixed effects, the random vector δ_1 is of order $q \times 1$ such that $\delta_1 \sim N\left(1, \Omega_g\right)$ and

$$\Omega_g = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1q} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2q} \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{q1} & \sigma_{q2} & \dots & \sigma_{qq} \end{bmatrix}$$

is the corresponding symmetric covariance matrix. From [1] and [2] the combined model for group j is given by

$$Y_j = X_j \beta + W_j \delta_j + \varepsilon_j \qquad (3)$$

Where, $X_j = \begin{bmatrix} W_j Z_j \end{bmatrix}$. By combining the data for all groups, we have

$$Y = X\beta + W\delta + \varepsilon \qquad (4)$$

Where, $Y = \left(Y_1^T, \dots, Y_m^T\right)^T$, $X = \left(X_1^T, \dots, X_m^T\right)^T$, W is the individual-level block diagonal matrix with W_j in the corresponding block, $\varepsilon = \left(\varepsilon_1^T, \dots, \varepsilon_m^T\right)^T$ and $\varepsilon \sim N\left(0, \sigma_e^2 I\right)$, $\delta \sim N\left(0, \Omega\right)$, and $\Omega = diag\left(\Omega_g\right)$ with Ω_g being the group-level covariance matrix. This implies that $E(Y) = X\beta$ and

$$Var(Y) = \Sigma = W\Omega W^T + \sigma_e^2 I \qquad (5)$$

In the next section, some estimation methods that are frequently used in practice and are also embedded in many software packages are highlighted. Focus is given to the estimation of both the model fixed effects and the variance components.

Estimation methods

Restricted maximum likelihood estimation: The first step of the analysis is to obtain estimates of the unknown fixed effects in β, as well as the estimates of the unknown variance components contained in Ω and σ_e^2, which are defined in the previous section. The common estimation methods for the multilevel regression model are ML and REML methods. ML estimators of the variance parameters are biased downwards, especially in small samples, since they do not take into account the degrees of freedom lost in the estimation of the fixed effects. The REML estimation method under model is presented as in the sequel. Unlike ML, REML maximizes the likelihood of the linearly independent error contrasts of linear combinations of the data Y, that are orthogonal to the explanatory design matrix. The linear combinations are chosen as $A^T Y$ instead of Y such that

$$A^T Y = A^T X \beta + A^T (W\delta + \varepsilon) \qquad (6)$$

Where, A is an idempotent matrix of rank $n - p$, and then $E(A^T Y) = 0$ if and only if $A^T X \beta = 0$. Since $Y \sim N(X\beta, \Sigma)$ and $A^T X \beta = 0$, then $A^T Y \sim N(0, A^T \Sigma A)$. Assuming $\text{cov}(\delta_{jh}, \delta_{jl}) = 0, h \neq l, \quad j = 1, \ldots, m, \quad h, l = 1, \ldots, q,$ the corresponding likelihood function to [6] can be expressed as

$$L(A^T Y) = (2\pi)^{-0.5(n-p)} | A^T \Sigma^{-1} A |^{-0.5} \exp\left(-0.5(A^T Y)^T (A^T \Sigma^{-1} A)^{-1} (A^T Y)\right)$$

Let, $P = A(A^T \Sigma^{-1} A)^{-1} A^T$ then after some algebra, the log-likelihood function is given by:

$$\log L(A^T Y) = \frac{-1}{2} \left\{ (n-p)\log(2\pi) + \log|\Sigma| + \log|X^T \Sigma^{-1} X| + Y^T P Y \right\} \quad (8)$$

Differentiating the log-likelihood function [8] with respect to σ_e^2 and $\sigma_r^2 (r = 1, \ldots q)$ yields

$$\frac{\partial \log L(A^T Y)}{\partial \sigma_e^2} = -\frac{n-p}{2\sigma_e^2} + \frac{Y^T P Y}{4\sigma_e^4} \qquad (9)$$

$$\frac{\partial \log L(A^T Y)}{\partial \sigma_r^2} = \frac{-1}{2} \left\{ tr(P W_r W_r^T) - Y^T P W_r W_r^T P Y \right\} \quad (10)$$

Set equal to zero and solve it we have a REML estimator for the residual variance as

$$\hat{\sigma}_e^2 = \frac{Y^T \hat{P} Y}{n - p} \qquad (11)$$

Where, $\hat{P} = A(A^T \hat{\Sigma}^{-1} A)^{-1} A^T$ that must be computed iteratively. By recalculate the new solutions of log-likelihood to obtain the new estimates of variance components and reformulate the new matrix $\hat{\Sigma}$. The process continues until convergence. REML estimation includes no procedure for estimating fixed effects. However, it would seem to use maximum likelihood estimation of β where $\hat{\beta} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} Y$ with $\hat{\Sigma}$ being the REML estimate of Σ (Searle et al. (1992), p. 254). One important assumption underlying REML estimation method is normality of the error distributions. When the residual errors are not normally distributed, the parameter estimates produced by the REML method are still asymptotically unbiased. However, the asymptotic standard errors are incorrect. Significance tests and confidence intervals can thus not be trusted (Goldstein, 2010). This problem does not completely vanish when the sample size gets larger.

REML sandwich estimator

One method to obtain better tests and confidence intervals is to correct the asymptotic standard errors of the fixed coefficients using the sandwich estimator which makes them less dependent on the distributional assumptions [8,9]. The method yields asymptotically consistent covariance matrix estimates without making distributional assumptions even if the assumed model is incorrectly specified with this respect. The sandwich estimator is often called the robust covariance matrix estimator. In the ML approach, the usual estimator of sampling variances and covariance is the inverse of the information matrix. Using matrix notation, the asymptotic covariance matrix of the estimated coefficients under REML can be written as follows

$$V_A = B^{-1} \qquad (12)$$

Where, B is the information matrix [10]. Then the sandwich estimator is given as

$$V_R = B^{-1} R B^{-1}$$

where the covariance matrix V_R is the robust covariance matrix of the regression coefficients, and R is the correction matrix that is based on the observed raw residuals. If the residuals follow a normal distribution, V_A and V_R are both consistent estimators of the covariances of the regression coefficients, but the model based asymptotic covariance matrix V_A is more efficient because it leads to the smallest standard errors. However, when the residuals do not follow a normal distribution, the model-based asymptotic covariance matrix is incorrect, while the sandwich estimator of V_R is still a consistent

estimator of the covariances of the regression coefficients. This makes the inference based on the robust standard errors less dependent on the assumption of normality.

Iterative generalized least square estimation

The IGLS method estimates regression coefficients, variances and their random effects. Under the assumption of residuals have multivariate normal, the method constructs an additional linear model whose unknown parameters are the between-group covariances σ_{jk} that represent the covariance between the jth and the kth element of α (i.e. the correlations between first-level parameter estimates), $j, k = 1, .., q$ and the within-group σ_e^2 variance.

Iterative generalized least square

This method begins with estimating the components of Ω and σ_e^2 in a linear modeling framework. We begin by expanding, re-expressing the combined covariance matrix Σ as a linear combination of the between-groups covariances σ_{jk} and the residual variances σ_e^2. This allows a subsequent formulation of a general linear model with the unknown parameters σ_{jk} and σ_e^2 being estimated. Let H be defined as a matrix that selects the element of the variance components such that $H_{j,k}$ is a $q \times q$ indicator matrix which is 0 in every element except the $(j,k)^{th}$ cell. Then, Σ can be written as

$$\Sigma = \sum_{j=1}^{q} \sum_{k=1}^{q} \sigma_{jk} W \left(I_m \otimes H_{j,k} \right) W^T + \sigma_e^2 I \qquad (14)$$

For a pictorial representation of see an interesting example in Lindquist et al. [11]. The next step in IGLS is to formulate the linear model that estimates the variance components. This is achieved based on by vectorizing the lower triangular and diagonal elements of the matrix corresponding to each unknown variance component. This can be done using the vech operator, which when applied to a matrix stacks its columns after removing all supra-diagonal elements. Using this notation, the summands in can be written $vech \left(W \left(I_m \otimes H_{j,k} \right) W^T \right)$ and $vech(I)$. These become the regressors in the new design matrix X^* (where * will be used to indicate the linear model for variance components, following the notation of Goldstein [6]). The response variable Y^* in this new model is based on the residual covariance matrix U where,

$$U = \left(Y - X \widehat{\beta}_{IGLS} \right) \left(Y - X \widehat{\beta}_{IGLS} \right)^T \qquad (15)$$

$$Y^* = vech(U) \qquad (16)$$

Y^* and X^* are used to estimate the variance components in $\beta^* = \{ \sigma_{jk}, \sigma_e^2 \}, (\text{for } j, k = 1, .., q)$ by using the following equation of the extra linear model

$$Y^* = X^* \beta^* + \varepsilon^* \qquad (17)$$

The IGLS procedure alternates between estimating β and $\beta*$ until convergence. The specific steps are as follows

A. Start with OLS estimates of the covariance, setting $\hat{\Sigma} = \sigma_e^2 I$.

B. Estimate fixed effects. Use the current estimate of $\hat{\Sigma}$ to calculate $\widehat{\beta}_{IGLS}$ by the GLS solution:

$$\widehat{\beta}_{IGLS} = (X^T \widehat{\Sigma}^{-1} X)^{-1} X^T \widehat{\Sigma}^{-1} Y \qquad (18)$$

C. Estimate variance components. Use the residuals to form Y^* and update $\widehat{\beta}^*_{IGLS}$ estimates for the within and between-groups variance components, again using the GLS procedure. Following Goldstein [6], the covariance of the variance components can be shown to be $\hat{\Sigma}^* = \hat{\Sigma} \otimes \hat{\Sigma}$ This result holds if and only if the data are multivariate normal or if the sample variance matrix is Wish art distributed. Thus, the GLS solution is given by

$$\widehat{\beta}^*_{IGLS} = \left(X^{*T} \hat{\Sigma}^{*-1} X^* \right)^{-1} \left(X^{*T} \hat{\Sigma}^{*-1} Y^* \right) \qquad (19)$$

If negative estimates are obtained for the variance components $\sigma_{jk}, j = k$ and σ_e^2 they are truncated to *0*. However, the covariance terms $\left(\sigma_{jk}, j \neq k \right)$ are allowed to be negative. Reform $\hat{\Sigma}$ from $\widehat{\beta}*$

4. Repeat steps 2–3 until convergence.

Sandwich IGLS

As in the ML case, we can obtain IGLS sandwich estimators that correct the asymptotic standard errors of the fixed parameters. Consider first the fixed part of the model and the usual IGLS estimate of the fixed parameters based upon the random parameter estimates using. The covariance matrix of these estimates is:

$$Cov(\widehat{\beta}_{IGLS}) = \left(X^T \hat{\Sigma}^{-1} X \right)^{-1} X^T \hat{\Sigma}^{-1} \{ Cov(Y) \} \hat{\Sigma}^{-1} X \left(X^T \hat{\Sigma}^{-1} X \right)^{-1} \quad (20)$$

Where, $Cov(Y) = \Sigma$ and is unknown. If we replace the central covariance term $Cov(Y)$ by the assumed model estimated value, $\hat{\Sigma},$ we obtain the usual formula

$$Cov(\hat{\beta}_{IGLS}) = (X^T \hat{\Sigma}^{-1} X)^{-1} \qquad (21)$$

The sandwich estimator is formed by replacing the estimate of central covariance term, $Cov(Y)$ in [21] by a robust estimator based on U in, and then the sandwich estimator for the parameter standard errors is constructed as

$$Cov(\hat{\beta}_{IGLS}) = \left(X^T \hat{\Sigma}^{-1} X \right)^{-1} X^T \hat{\Sigma}^{-1} U \hat{\Sigma}^{-1} X \left(X^T \hat{\Sigma}^{-1} X \right)^{-1} \quad (22)$$

Further details of the Huber-White correction for the multilevel model are given by Goldstein and Raudenbush & Bryk.

Expectation maximization

The EM algorithm addresses the problem of maximization likelihood as a problem of missing data.

EM algorithm: Each EM step requires sums of squares and product of the conditional means of the random effects (given the data and current parameter estimates) as well as the conditional variances of these random effects. Consider the combined model in [10]. Then, the alogritm begins with the M-step followed by the E-step as shown in the sequel.

M step: The EM algorithm conceives Y_j as the observed data with δ_j as the missing data. Thus, the complete data set is (Y_j, δ_j). To estimate β, this is achieved by deducting $W_j \delta_j$ from both sides of [12], yielding

$$Y_j - W_j \delta_j = W_j Z_j \beta + \varepsilon_j \qquad (23)$$

and justifying the OLS estimate as the complete-data ML estimate of βfor all groups. Thus, considering

$$\hat{\beta} = \left(\sum_j (W_j Z_j)^T (W_j Z_j) \right)^{-1} \sum_j (W_j Z_j)^T (Y_j - W_j \delta_j) \qquad (24)$$

The complete-data ML estimates of Ω and σ_e^2 are similarly straightforward, where

$$\hat{\Omega} = m^{-1} \sum_j \delta_j \delta_j^T \qquad (25)$$

$$\hat{\sigma}_e^2 = n^{-1} \sum_j \hat{\varepsilon}_j^T \hat{\varepsilon}_j^T \qquad (26)$$

And,

$$\hat{\varepsilon}_j = Y_j - W_j Z_j \hat{\beta} - W_j \delta_j$$

E step: The CDSS are not observed, but they can be estimated by their conditional expression, given the data Y_j and the parameter estimates from the previous iteration. Dempster et al. showed that substituting the expressed CDSS for the M-step formulas would produce new parameter estimates having higher likelihood than the current estimates. Consider multivariate normal distribution data in the form

$$\begin{pmatrix} Y_j \\ \delta_j \end{pmatrix} \sim N \left(\begin{pmatrix} W_j Z_j \beta \\ 0 \end{pmatrix}, \begin{pmatrix} W_j \Omega_g W_j^T + \sigma_e^2 I_{n_j} & W_j \Omega_g \\ \Omega_g W_j^T & \Omega_g \end{pmatrix} \right) \quad (27)$$

Then,

$$\delta_j \mid Y_j, \beta, \Omega_g, \sigma_e^2 \sim N(\delta_j^*, \sigma_e^2 c_j^{-1})$$

$$\delta_j^* = c_j^{-1} W_j^T \left(Y_j - W_j Z_j \beta \right)$$

$$c_j^T = W_j^T W_j + \sigma_e^2 \Omega_g^{-1} \qquad (29)$$

Now, having identified the CDSS needed for M step, the EM algorithm is readily specified as;

I. Estimate the CDSS data:

$$E \left(\sum_j (W_j Z_j)^T W_j \delta_j \mid Y_j, \beta, \Omega, \sigma_e^2 \right) = \sum_j (W_j Z_j)^T W_j \delta_j^*$$

$$E \left(\sum_j \delta_j^* \delta_j^{*T} \mid Y_j, \beta, \Omega, \sigma_e^2 \right) = \sum_j \delta_j^* \delta_j^{*T} + \sigma_e^2 \sum_j c_j^{-1}$$

$$E \left(\sum_j \delta_j^* \delta_j^{*T} \mid Y_j, \beta, \Omega, \sigma_e^2 \right) = \sum_j \delta_j^* \delta_j^{*T} + \sigma_e^2 \sum_j c_j^{-1}$$

$$E \left(\sum_j \varepsilon_j^T \varepsilon_j \mid Y_j, \beta, \Omega, \sigma_e^2 \right) = \sum_j \varepsilon_j^{*T} \varepsilon_j^* + \sigma_e^2 tr \sum_j c_j^{-1} W_j^T W_j$$

Where,

$$\varepsilon_j^* = Y_j - W_j Z_j \hat{\beta} - W_j \delta_j^*$$

II. Substitute the estimated CDSS in to the M-step formulas to obtain new estimates of the parameters.

III. Feed the new parameter estimate into step 1.

IV. Continue until changes in the log-likelihood become sufficiently small or the largest change in the value of the parameters is sufficiently small.

Measures of assessment

Beside the review of some estimation methods, one may be interested in knowing which method may be more useful when the standard multilevel model assumptions are not met in the available data. In this section, we provide some notes about three commonly used measures by which one could assess the estimated parameters and compare their performance. This shall guide the practitioners to use the better method whenever their data reveal the need for special inferential procedures.

Percentage relative bias

The accuracy of the parameter estimates can be calculated using the mean relative bias. Let $\hat{\theta}$ represent the estimate of the population parameter θ, then the percentage relative bias is given by

$$\text{Relative bias} = \frac{\hat{\theta}}{\theta} \times 100 \qquad (30)$$

The parameter θ and its estimator $\hat{\theta}$ can be any regression coefficient that we considered in and the choices among the ML estimator or the estimators in and for fixed effects as well as for the variance components under the corresponding estimation method. A better estimator could be checked by testing whether its relative bias differs from one. With α=0.001, the p-values of the estimated parameters are Bonferroni-corrected [12]. The Bonferroni correction issued to adjust hypothesis being tested and the confidence intervals by testing each individual hypothesis at a significance level of $\alpha / \tilde{n}$, where α is the desired overall alpha level and $\tilde{n}$ is the number of hypotheses. Mass & Hox [13] provided a simulation study to assess the accuracy of the parameter estimates using the percentage relative bias when the normality assumption of the error distribution is violated. The study showed that, the percentage relative bias is the same for the REML and robust REML, since it focuses on investigate the parameter estimates and not their standard errors. The study resulted in only one significant difference when the group-level residuals were generated from chi-square distribution as the worst condition where the number of groups was 30, group size equal 5 and interclass correlation coefficient equal 0.1. In a different context, Korendijk et al. [14] highlighted the effect of ignoring the group-level hetroscedasticity on the estimation of regression coefficients, the variances components, and their standard errors. The authors provided a simulation study using the percentage relative bias and coverage probability (see the next paragraph).The results showed that the individual-level regression coefficients, variances, and their standard errors are unbiased. However, the standard errors of group-level variances are underestimated.

Confidence intervals

To assess the accuracy of the standard errors, another useful measure is the coverage probabilities of estimation. For each parameter set a 95% confidence interval using the asymptotic standard normal distribution of the estimator. First, we obtain the actual coverage probability, which is counted as the proportion of the overall iterations in which the confidence interval covers the true values of true parameters. We are concerned with how far the actual coverage probability falls below the nominal level. Using the null hypothesis; H_0 : Actual Coverage probability $= 1 - \alpha$ the deviation is analyzed with a simple z-score, which is calculated based on normal approximation to the binomial distribution as follows

$$Z = \frac{C - 1000(1-\alpha)}{\sqrt{1000(1-\alpha)\alpha}} \qquad (31)$$

Where, c is the count that θ is covered by the confidence interval, 1000 is the number of replicates in each condition and $(1-\alpha)$ is the intended coverage probability. The downward

with this method appear when it used for estimated variances of random effects since it doesn't take their asymptotic distributions into account. For this reason, we can depend on asymptotic relative efficiency to assess the accuracy of the variance components in multilevel models. In a comparison between the REML as a parametric method and the robust REML as an alternative robust parametric methods, Mass and Hox [13] introduced a simulation study using coverage probability for both fixed and random parameters in the model to evaluate the accuracy of the parameter estimates and their standard errors when the assumptions of normality and large samples are not met. A variety of factors were considered including the number of groups, group size, interclass correlation coefficient and the group-level residuals generated from three non-normal (uniform, Laplace and chi-square)distributions. The study shows that under the violation of normality assumption at the group-level residuals, there is no effect of such factors on the parameter estimates. On the other hand, for the random part in the model the standard errors of variances at the group-level are highly inaccurate and thus the robust REML do better than REML especially in large number of groups.

In dealing with the problem of determining the effect of different sample sizes at the group-level on the accuracy of the estimates in multilevel regression models, Mass & Hox [15] provided a simulation study based on the aforementioned coverage probability measure to determine the effect of number of number of groups, group sizes and interclass correlation coefficients on the parameter estimates and their standard errors. The results show that having small number of groups leads to biased estimates of the group-level standard errors of the regression coefficients. However, for the remaining factors the estimates of the parameters (regression coefficients and variance components) and their standard errors are unbiased and yet accurate.

Asymptotic relative efficiency

To assess the performance of the variance components, the asymptotic relative efficiency (ARE) is defined as one of the principal comparison measures. The ARE of two procedures is the ratio of their efficiencies measures where efficiencies are often defined using the asymptotic mean squared error as one of the measures of desirability. In order to study the sample properties of two estimators under a simulation study, the ARE can be calculated according to the following steps:

o For each case study we generate, say 1000, simulated data sets.

o Let θ be the target population parameter and $\hat{\theta}_{hi}$ denote the value of the estimator for the first method in the ith simulation, and $h = 1, 2$.

o Denote the MSE for the first estimation method, that,

$$MSE_h = \frac{1}{1000} \sum_{i=1}^{1000} \left(\left(\hat{\theta}_{hi} - \theta \right)^2 \right) \qquad (32)$$

Where, the MSE_h is a measure of the efficiency of the hth estimator and values closer to zero are more preferred.

o Repeat the previous steps for each method of estimation and compare between each method using the ARE criterion

$$ARE = \frac{MSE_1}{MSE_2} \qquad (33)$$

Based on Mckean et al. [16] investigate the robust approach (weighted rank based analyses of linear models) introduced by Hettmansperger & Mckean [17]. The study compared the ARE of the weighted Wilcoxon and OLS estimators in terms of their asymptotic variances and concluded that under normal errors, the Wilcoxon estimator is 95% as efficient as the OLS method. Thus, there is only a 5% loss in efficiency if the Wilcoxon estimator is used and the error distribution is actually normal. However, the ARE is usually greater than 1 if the true distribution has tails heavier than those under a normal distribution (e.g. contaminated normal distribution or the data corrupted by outliers).

Under linear mixed models, which represent the general case of multilevel regression models, Kloke et al. [18] extended the robust approach introduced by Hettmansperger and Mckean [16] to linear mixed models with covariates using general score functions. The authors compared between the OLS and weighted Wilcoxon using the ARE as an assessment measure and concluded that the robust analysis is more effective than OLS in the presence of outliers or underlying error distributions with heavy tails. Besides, Mckean & Kloke [19] proposed a family of optimal score functions under contaminated normal distributions of the error terms in both linear and nonlinear models [20-25]. In this study they compared Wilcoxon, OLS, and ML in terms of their asymptotic variances by using the ARE to conclude about the efficiency and validity of Wilcoxon over skewed-normal and contaminated normal distributions [26-29].

Conclusion

In this article, some estimation methods for multilevel models are reviewed. In addition, some measures that assess the behavior, reliability, and efficiency of those estimators are highlighted [30-32]. One important remark is that the efficiencies of the variance components estimators that are produced under each estimation method may not be easily compared using the well-known coverage probability measure. The ARE becomes a useful candidate in this case. A future research point of assessing the behavior of the fixed effects and the variance components under multilevel models with more than two levels are complicated covariance structures is needed to choose among the competing estimation methods and their robust versions. The same recommendation shall apply to linear mixed models.

References

1. Snijders TAB, Bosker RJ (1999) An Introduction to basic and advanced multilevel modeling. London.

2. Searle SR, Casella G, McCulloch CE (1992) Variance Components, JohnWiley and Sons, New York, USA.

3. Rao CR (1971) Estimation of variance and covariance components -MINQUE theory. J Multivariate Anal 1: 257-275.

4. Hox JJ (2002) Multilevel analysis: Techniques and applications. Mahwah, NJ: Erlbaum.

5. Bryk AS, Raudenbush SW (1992) Hierarchical Linear Models. Sage, Newbury Park, CA.

6. Goldstein H (1986) Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least-Squares. Biometrika 73: 43-56.

7. Goldstein H (1995) Multilevel Statistical Models .London: Institute of Education.

8. Huber P J (1967) The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions, University of California Press, Berkely, 221-233.

9. White H (1982) Maximum likelihood estimation of misspecified models, Econometrics. 50: 1-50.

10. Eliason SR (1993) Maximum Likelihood Estimation: Logic and Pract, SAGE University Papers.

11. Lindquist MA, Spicer J, Asllani I, Wager TD (2012) Estimating and Testing Variance Components in a Multilevel GIM, NeuroImage 59: 490-501.

12. Dunn, Olive Jean (1961) Multiple Comparisons Among Means. Journal of the American Statistical Association. 56 (293): 52-64

13. Mass J MC, Hox JJ (2003) The Influence of Violations of Assumptions on Multilevel Parameter Estimates and Their Standard Errors. Computational statistics & Data analysis 46: 427-440.

14. Korendijk EJH, Mass JMC, Mirjam Moerbeek, Van der Heijden PGM (2008) The Influence of Misspecification of the Hetroscedasicity on Multilevel Regression Parameters and Standard Error Estimates, Methodology4(1): 67-72

15. Mass JMC, Hox JJ (2005) Sufficient Sample Sizes for Multilevel Modeling. Methodology 1(3): 86-92.

16. Mckean JW, Naranjo JD, Sheather SJ (1999) Diagnostics for comparing robust and least squares fits. Journal of Non-parametric Statistics 11: 161-188.

17. Hettmansperger TP, Mckean JW (1978) A Robust Analysis of the General Linear Models Based on One Step R-Estimate. Biometrika 65: 571-579.

18. Kloke JD, Mckean JW, Rashid M (2009) Rank-Based Estimation and Associated Inferences for Linear Models with Cluster Correlated Errors. Journal of the American Statistical Association 104: 384-390.

19. Mckean JW, Kloke JD (2014) Efficient and adaptive rank-based fits for linear models with skewed-normal errors, Journal of Statistical Distributions and Applications 1: 1-18.

20. Anderson RL, Bancroft T A (1952) Statistical Theory in Research. McGraw-Hill, New York.

21. Goldstein H (1988) A General Model for The Analysis of Multilevel Data. Biometrika 53(4): 455-467.

22. Goldstein H (1989) Restricted Unbiased Iterative Generalized Least-Square Estimation. Biometrika, 76(3): 662- 663.

23. Goldstein H, Rasbash J (1992) Efficient Computational Procedures for The Estimation of Parameters in Multilevel Models Based on Iterative Generalized Least Square. Computational Statistics & Data Analysis 13: 63- 71.

24. Harville DA (1977) Maximum Likelihood Approaches to Variance Component Estimation and Related Problems. Journal of the American Statistical Association 72(358): 320-338.

25. Henderson CR (1950) Estimation of genetic parameters. Ann Math Stat 21: 309-310.

26. Henderson CR (1953) Estimation of variance and covariance components. Biometrics 9: 226-252.

27. Henderson CR, Kempthorne O, Searle SR, Von Krosig CN (1959) Estimation of environmental and genetic trends from records subject to culling. Biometrics 15: 192-218.

28. Hettmansperger TP, Mckean JW (1998) Robust Nonparametric Statistical Methods, Arnold, London.

29. Mckean J W (2004) Robust Analysis of Linear Models, Statistical Science 19: 562-570.

30. Mckean JW, Terpstra JW, Kloke JD (2009) Computational rank-based statistics. Wiley Interdisciplinary Reviews: Computational statistics 1: 132-140.

31. Patterson HD, Thompson WA (1971) Recovery of Inter-Block Information When Block Sizes are Unequal. Biometrica 58(3): 545-554.

32. Searle SR (1982) Matrix Algebra Useful for Statistics, JohnWiley and Sons, New York.

Your next submission with Juniper Publishers
will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php