A Monte Carlo Simulation Study of the Properties of Residual Maximum Likelihood (REML) Estimators for the Linear Gaussian Mixed Model

K. Nadarajah, M.R. Sooriyarachchi

Department of Statistics, University of Colombo

Contact Person: **Prof. Roshini Sooriyarachchi,**

Department of Statistics,

University of Colombo,

Colombo 03.

Telephone: **2590111**

Fax: **2590111**

E-mail: **roshini@mail.cmb.ac.lk**

Abstract

The linear Gaussian mixed model is a tool box for analyzing experimental as well as non experimental designs in a flexible way elaborately. It is the model that contains mixtures of fixed effects as well as random effects. There are several ways to estimate fixed effects and variance components of the random effects. The most commonly used methods are Iterative Generalized Least Squares (IGLS) Estimation, Maximum Likelihood (ML) Estimation and Residual Maximum Likelihood (REML) Estimation. Of these methods many researchers prefer the REML method. This method is an iterative method thus its properties cannot be studied analytically. In the past simulation studies have been used only to study the properties of unbiasedness and efficiency of these REML estimators. These simulation studies have been of a small scale and usually have examined estimation of only either fixed or random effects but not both. Also the affect of varying sample size on the properties of the estimators have not been studied. Therefore the aim of this paper is to study the major desirable properties of estimators, namely, unbiasedness, consistency, sufficiency and efficiency for the REML method of estimation for both fixed and random effects for varying sample sizes and for varying ratios of variance of random effect to error variance. This was achieved by using an extensive Monte Carlo Simulation study. Code for this simulation study was developed using Java programming language. The results indicate that the Residual Maximum Likelihood estimation (REML) method holds all the desired properties for fixed effects. However for variance components of random effects and errors it does not hold the property of sufficiency and also though when the ratio of variance of random effects to error variance is small it holds the property of efficiency it is not so efficient when this ratio is large.

Keywords : Residual Maximum Likelihood (REML) Estimation, Mixed Models, Monte Carlo Simulation, Properties of Estimators, Minimum Variance Quadratic Unbiased Estimators (MIVQUE)

1.Introduction

The linear Gaussian mixed model (Brown and Prescott, 2001) is widely used in a number of areas including Biology, Medicine and Social Sciences. This model encompasses methodology to relax the assumption of independence and variance homogeneity. The key distinguishing feature of mixed models compared with fixed effects models is that they are able to model data in which the observations are not independent. A Mixed model consists of fixed (non-random) effects as well as random effects. Each random effect in the model gives rise to a variance component. The theory behind mixed models and the advantages of mixed models over the traditional fixed effects models have been discussed by Littell et al. (1996), Verbeke and Molenberghs (1997,2000), and Brown and Prescott (2001). There are several ways to estimate fixed effects and variance components of the random effects such as Maximum Likelihood Estimation (ML), Residual Maximum Likelihood Estimation (REML), Iterative Generalized Least Squares Estimation (IGLS) and Least Squares Estimation (LSE). Hartley and Rao 1967; Patterson and Thompson 1971; Harville 1977; Laird and Ware 1982; Jennrich and Schluchter 1986 have found that in many situations, the best approach for the estimation of fixed effects and random components is to use likelihood-based methods, exploiting the assumption that the random components are normally distributed. Of the two commonly used likelihood-based methods, maximum likelihood (ML) and restricted/residual maximum likelihood (REML), REML has been found to be the more desirable. As REML estimation is based on an iterative process the properties of the REML estimators cannot be studied analytically thus in the past researchers such as Swallow and Monahan (1984), McGilchrist (1988), Zhang, Zhang, Liu, Haussmann (1995), Kenward, Roger (1997) have used Monte-Carlo simulation for this purpose. All these studies have been comparative studies where REML estimators have been compared with other estimators and the authors have only examined the properties of unbiasedness or/and efficiency in estimators of fixed or random effects. Also no study has been carried out to examine the effect of sample size on these estimators.

The superiority of REML estimators has already been established by the previously mentioned simulation studies. Thus in this paper the objective was to examine the properties unbiasedness, efficiency, sufficiency and consistency of REML estimators of both fixed and random effects in the general linear Gaussian mixed model by using Monte Carlo simulation over a range of sample sizes. In this simulation study balance data is considered and special forms of variance covariance matrices for the errors and random effects are assumed. To achieve this objective Java programs were developed by the authors.

Section 2 consists of methodology and comes in three parts. The first part explains the REML estimation procedure for the estimation of fixed effects and variance components of the random effects, the second part discusses the properties of unbiasedness, efficiency, sufficiency and consistency of an estimator and the third part illustrates the simulation procedure used in the study. Section 3 consists of the results of the simulation study and the conclusions drawn from these results. A discussion of the findings is given in section 4.

2. Methodology

Section 2.1 describes the REML estimation method for mixed models, section 2.2 defines the main desirable properties of estimators and section 2.3 illustrates the simulation procedure. In order to set the stage for the description of material for sections 2.1-2.3 some notation and definitions are necessary and are given below.

The general linear Gaussian mixed model is defined as : $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$ **w**here, $\mathbf{y}$ is an $n \times 1$ vector of observations, $\mathbf{X}$ is an $n \times p$ matrix of known constants of rank p, $\boldsymbol{\alpha}$ is a $p \times 1$ vector of fixed effects parameters, $\mathbf{Z}$ is an $n \times r$ matrix of known constants of rank r and $\mathbf{b}$ is a $r \times 1$ vector of random variables. It is assumed that the errors $\boldsymbol{\varepsilon}$ follow a normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{R}$ ($\boldsymbol{\varepsilon} \sim N(\mathbf{0},\mathbf{R})$), the random effects $\mathbf{b}$ follow a normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{G}$ ($\mathbf{b} \sim N(\mathbf{0},\mathbf{G})$) and $\boldsymbol{\varepsilon}$ and $\mathbf{b}$ are uncorrelated. This implies that the mean of $\mathbf{y}$ is $\mathbf{X}\boldsymbol{\alpha}$ and $\mathbf{V}$, the variance of $\mathbf{y}$, is $\mathbf{ZGZ'} + \mathbf{R}$.

2.1 REML method of estimation of fixed effects and variance components in Mixed models

In this approach (Liao and Lipsitz, 2002), the parameter α is eliminated from the log likelihood so that it is defined only in terms of the variance parameters. Here, the likelihood function is obtained based on the residual terms which are known as full residuals, $y - X\hat{\alpha}$. Therefore the joint log likelihood for α and the variance parameter, γ, can be expressed as a product of the likelihoods based on $y - X\hat{\alpha}$ and $\hat{\alpha}$:

$$L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y}) = L(\boldsymbol{\gamma}; \mathbf{y} - X\hat{\alpha})\, L(\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}, \boldsymbol{\gamma}) \quad \ldots\ldots\ldots(1)$$

. Thus, the likelihood for γ based on $\mathbf{y} - X\hat{\alpha}$ is given by :

$$L(\boldsymbol{\gamma}; \mathbf{y} - X\hat{\alpha}) = L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y}) / L(\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}, \boldsymbol{\gamma}) \text{........ (2)}$$

For fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ the likelihood of $\mathbf{y}$, $L(\boldsymbol{\gamma}, \boldsymbol{\alpha}; \mathbf{y}) \propto |V|^{-1/2}\, e^{-\frac{1}{2}(Y-X\alpha)'V^{-1}(Y-X\alpha)}$ ....(3)

Equation (3) gives the maximum likelihood estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ to be $\hat{\alpha} = (X'\hat{V}^{-1}X)^{-}X'\hat{V}^{-1}y$ and $\hat{\gamma} = \hat{G}Z'\hat{V}^{-1}(y - X\hat{\alpha})$.

Assuming V is known, $\hat{\alpha}$ has a multivariate normal distribution with mean $\boldsymbol{\alpha}$ and variance $(X'V^{-1}X)^{-1}$. Hence, $L(\alpha;\hat{\alpha};\gamma) \propto |X'V^{-1}X|^{-1/2}\, e^{-\frac{1}{2}((\hat{\alpha}-\alpha)'X'V^{-1}X(\hat{\alpha}-\alpha))}$

This can be reduced to :

$$L(\alpha;\hat{\alpha};\gamma) \propto |X'V^{-1}X|^{-1/2}\, e^{-\frac{1}{2}((y-X\alpha)'V^{-1}(X\hat{\alpha}-X\alpha))} \text{................(4)}$$

Substituting equations (3) and (4) in (2) and taking $\alpha = \hat{\alpha}$ the residual maximum likelihood is obtained as $L(\gamma; y - X\hat{\alpha}) \propto |V|^{-\frac{1}{2}} |X'V^{-1}X|^{\frac{1}{2}}\, e^{-\frac{1}{2}[(y-X\hat{\alpha})'V^{-1}(y-X\hat{\alpha})]}$ and thus the residual maximum log likelihood is

$$\log L(\gamma; y - X\hat{\alpha}) = k - \frac{1}{2}[\log|V| - |X'V^{-1}X| + (y - X\hat{\alpha})'V^{-1}(y - X\hat{\alpha})] \text{............(5)}$$

The difference between the residual maximum log likelihood and the ordinary log likelihood is caused by an extra term $\log|X'V^{-1}X|^{-1}$ which is the log of the determinant of var($\hat{\alpha}$). The residual maximum likelihood is equivalent to having integrated α out of the likelihood for α and γ, and for this reason residual maximum likelihood is sometimes referred to as a ‘marginal’ method. Because, the residual maximum likelihood takes account of the fact that α is a parameter and not a constant, the resulting variance parameter estimates are unbiased. As with maximum likelihood, α is then estimated by treating the variance parameters as fixed, and finding the values of α which maximize the residual maximum log likelihood.

Maximum likelihood and Residual Maximum likelihood methods work by obtaining variance parameter estimates that maximize a likelihood function. A solution cannot be specified by a single equation as it was for the fixed and random effects because the derivatives of the log likelihood with respect to the variance parameters are non-linear. An iterative process such as the widely applied Newton- Raphson algorithm is therefore required. This works by repeatedly solving a quadratic approximation to the log likelihood function. Although the function will not, in general, be quadratic, in the region of the Maximum likelihood solution the quadratic approximation is usually quite good, and the Newton- Raphson iterative procedure will usually converge appropriately. The iterative process can be defined by $\theta_{i+1} = \theta_i - f''^{-1}(\theta_i) f'(\theta_i)$

where $f'(\theta_i)$ and $f''(\theta_i)$ are the actual derivatives of $f(\theta_i)$ evaluated at θi. The matrix of second derivatives, $f''(\theta_i)$ is often referred to as the Hessian matrix. In mixed models f(θ) is taken to be the log likelihood expressed in terms of the variance parameters. The need to evaluate the derivatives at each iteration can make the Newton- Raphson algorithm computationally intensive. Computation can be made easier by using a matrix known as the information matrix in place of $f''(\theta_i)$ in the iterative process. The information matrix is the expected value of the Hessian matrix and it is easier to compute than the Hessian matrix because some of the correlation terms are zero. When it is used the process can be referred to as the method of scoring or Fisher scoring. This method has been shown to be more robust to poor starting values than the Newton- Raphson algorithm. An indication of the precision of the variance parameters can be obtained from an estimate of their variance and their degree of correlation from the covariances. However, this estimate is based on standard asymptotic (large sample) theory. The asymptotic covariances of the variance parameters are given by the negative of the expected values of second partial derivatives of the log likelihood. Since the resulting covariances are based on asymptotic theory and are related to the estimated variance parameter values themselves, they should be interpreted cautiously.

2.2 Desirable Properties of Estimators

There are several desirable properties that an estimator of a population parameter can possess. These properties are described in Mood, Graybill and Boes (1963) and are explained in this section.

1. Unbiasedness

Suppose $\hat{\theta}$ is an estimator of parameter θ. Then the bias of this estimator is defined to be Bias ($\hat{\theta}$) = E[$\hat{\theta}$]-θ= E[$\hat{\theta}$-θ]. An estimator is said to be unbiased if its bias is equal to zero for all values of parameter θ. *This property intuitively means that on average, the estimator will be equivalent to the population parameter.*

2. Consistency

Let $X_1, X_2, \ldots$ be a sequence of identical and independent random variables with common density function $F_x(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. A sequence of point estimators $T_n(X_1, X_2, \ldots, X_n) = T_n$ will be called consistent for $\Psi(\boldsymbol{\theta})$ if

$$T_n \xrightarrow{P} \Psi(\boldsymbol{\theta}) \quad \text{as } n \to \infty \text{ for each fixed } \boldsymbol{\theta} \in \Theta.$$

Remark: $T_n \xrightarrow{P} \Psi(\boldsymbol{\theta})$ if and only if $P\{|T_n - \Psi(\boldsymbol{\theta})| > \varepsilon\} \to 0$ as $n \to \infty$ for every $\varepsilon > 0$.

Property of consistency intuitively means that estimators taken far enough in the sequence are more likely to be in the vicinity of the parameter being estimated, and in the limit they will be arbitrarily close to $\boldsymbol{\theta}$ with probability one.

Mean Square Consistency

If $\lim_{n\to\infty} E[T_n(X) - \theta]^2 = 0$, then the sequence $T_n(X)$, the estimator for θ, is said to be consistent in quadratic mean.

Result

Consistent in quadratic in mean implies consistency of an estimator in general. But, it is not necessarily other way around.

3.Sufficiency

Let $X = (X_1, X_2, \ldots, X_n)$ be a random sample from $\{ F_\theta: \theta \in \Theta \}$. A statistic $T = T(X)$ is sufficient for θ or for the family of distributions $\{ F_\theta: \theta \in \Theta \}$ if and only if the conditional distribution of X, given T = t, does not depend on θ (except perhaps for a null set A, $P_\theta\{ T \in A \} = 0$ for all θ).

$Pr (X = x \mid T(X) = t, \theta) = Pr (X = x \mid T(X) = t)$, Or in shorthand $Pr (x \mid t, \theta) = Pr (x \mid t)$

A statistic is sufficient for a family of probability distributions if the sample from which it is calculated gives no additional information than does that statistic, as to which of those probability distributions is that of the population from which the sample was taken.

It is often hard to verify or disprove sufficiency of a statistic directly because we need to find the distribution of the sufficient statistic. The following theorem is often helpful.

Factorization Theorem

Let $X_1, X_2, \ldots, X_n$ denote the a random sample of size n from a distribution that has probability distribution function or probability mass function $f(x;\theta)$, $\theta \in \Omega$. A statistic $Y = u(X_1, X_2, \ldots, X_n)$ is sufficient for θ if and only if we can find two nonnegative functions, k_1 and k_2 such that for all sample points $(X_1, X_2, \ldots, X_n)$,

$$f(x_1;\theta) \ldots f(x_n;\theta) = k1[u(x_1, x_2, \ldots, x_n); \theta]\, k_2(x_1, x_2, \ldots, x_n)$$

where $k_2(x_1, x_2, \ldots, x_n)$ does not depend upon θ.

4.Efficiency

Assume that the regularity conditions of the FCR inequality are satisfied by the family of density functions $\{ F_\theta: \theta \in \Theta \}$, $\Theta \subset \Re$. We say that an unbiased estimator T for parameter θ is most efficient for the family $\{ F_\theta \}$ if

$$\text{var}_\theta\,(T) = \left\{E_\theta\left[\frac{\partial \log f_\theta(X)}{\partial\theta}\right]^2\right\}^{-1} = \text{In}(\theta).$$

The relative efficiency of two procedures is the ratio of their efficiencies, although often this term is used where the comparison is made between a given procedure and a notional "best possible" procedure. Efficiencies are often defined using the variance or mean square error as the measure of desirability.

In this study the efficiency of the REML estimators relative to the Minimum Variance Quadratic Unbiased Estimators (MIVQUE) are examined for the random components.

MIVQUE and its variance

Swallow and Searle (1978) summarize Rao's (1971b) derivation of MIVQUE. To compute the MIVQUE's the user must supply a priori values for the variance components. The estimators are then functions of the data and of the a priori values, and are only locally minimum variance; that is, they are minimum variance only when each a priori value equals the true value of the corresponding variance component. Realistically, the user cannot provide perfect a priori values, so, in application, the estimators will not be minimum variance. Equations (17) and (18) of Swallow and Searle (1978) explains how to determine the variance of the MIVQUE.

2.3 Simulation Study for Properties of REML Estimators in Mixed Models

In order to study the properties of Residual Maximum Likelihood (REML) estimators a simulation study was used. In this simulation study the model considered is $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e}$ where $\mathbf{e} \sim N(\mathbf{0}, \sigma_e^2\mathbf{I})$ and $\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_b^2\mathbf{I})$ and $\mathbf{e}$ and $\boldsymbol{\beta}$ are uncorrelated**.** In this simulation study balance data is considered. The special case of $\mathbf{R}=\sigma_e^2\mathbf{I}$ and $\mathbf{G}= \sigma_b^2\mathbf{I}$ is assumed.

Here $\boldsymbol{\alpha}$, σ_b^2 and σ_e^2 are the parameters to be estimated through simulation. The fixed effect component was taken as $\boldsymbol{\alpha}' = (1\ 2\ 3\ 4)$. Several values of σ_b^2 (0.0, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0) were considered and σ_e^2 was fixed at 1.0. The sample sizes of 20, 60, 100, and 200 were examined. For each combination 10,000 simulation runs were made. Errors and random effect components were generated from normal distributions with corresponding parameters. According to the model defined in this study, design matrices were set.

Using the simulation study, estimates of fixed effects components, bias values, its true and estimated variance and estimates of variance components of random effects, bias values, and its true and estimated variance are calculated for each combination. The coding was done using Java and is available on request.

3. Results and Conclusions

Table 1 gives for the fixed effects (α), the parameter estimate ($\hat{\alpha}$), Bias (B($\hat{\alpha}$)), estimated variance (EV($\hat{\alpha}$)) and percentage difference in variance (P($\hat{\alpha}$)) obtained under REML method of estimation, for different values of sample size (n).

Table 1 should come here

The results in table 1 illustrate the following.

When the sample size and the fixed effect is held constant the parameter estimates of the fixed effects do not depend on the values of σ_b^2. This implies that, the assumption of "random effects being independent of error terms" is valid.

The bias values are between -0.015 to 0.02 for all the values σ_b^2 for sample size of 20. It is obvious from the bias values that the fixed effect components are unbiased even for small sample sizes. As sample size increases, bias values lie between -0.03 to 0.03. This implies that irrespective of the sample size, the fixed effects component parameter estimator is unbiased for the population parameter of fixed effects component.

In order to verify the property of consistency, variance estimates are needed. Table 1 gives the estimated variance of fixed effects components. According to table 1, estimated variance reduces and tends to zero as sample size increases for all the fixed effects components and for all the values σ_b^2. Therefore it can be concluded that the fixed effect component estimator is consistent for the population parameter α .

Sufficiency is another property to be looked at. For that percentage difference between estimated and true variance of fixed effects components estimates is needed. Table 1 gives the percentage values of difference between estimated and true variance of fixed effects components estimates in REML. According to table 1, the percentage of difference between the true and estimated variance estimates are between -6% and 6% for all the sample sizes. For small σ_b^2 values, the true and estimated variance estimates are quite similar. As σ_b^2 value increases, the percentage of difference between the true and estimated variance estimates also increase. But that increment does not affect the property of sufficiency as those values can be assumed as small. That is, the sample explains the population well for fixed effects. Therefore it can be concluded that the fixed effect component estimator is sufficient for the population parameter α .

When $\boldsymbol{G}$ and $\boldsymbol{R}$ are known (when **V** is known), Searle (1971), Harville : (1988), (1990), Robinson (1991) ,McLean, Sanders, and Stroup (1991) have shown that $\hat{\alpha}$ is the *best linear unbiased estimator* (BLUE) of α . Here, "best" means minimum mean squared error. However in practice **V** is rarely known. Further simulations carried out but not reported in this paper indicate that the REML estimator of α is more efficient than its corresponding Iterative Generalized Least Squares (IGLS) estimator and the Maximum Likelihood estimator.

Table 2 gives for the random effects, the parameter estimates, Bias, estimated variance(EV), percentage difference in variance(PDV) obtained under REML method of estimation and ratio of MIVQUE lower bound to the variance of REML estimator (Ratio) obtained for the random effects for different combinations of sample size (n).

Table 2 should come here

The results in table 2 illustrate the following.

With respect to table 2, bias values are very small for all the values of σ_b^2 for sample size of 20. It is obvious from the bias values that the random effect components are unbiased even for small sample sizes. As sample size increases, bias values decrease and they tend to zero. That implies that irrespective of the sample size, the random effects components parameter estimators $\hat{\sigma}_b^2$ and $\hat{\sigma}_e^2$ are unbiased for the population parameters σ_b^2 *and* σ_e^2 respectively.

In order to verify the property of consistency, variance estimates are needed. Table 2 gives the estimated variance of random effects components. According to table 2, though the estimates of the variance of variance component of random effects decrease with sample size, they do tend to zero for small σ_b^2. But for higher values of σ_b^2, variance of variance component of random effects decreases. With this nature of variance of variance component of random effects, it can be concluded that the estimator $\hat{\sigma}_b^2$ of the variance component of errors is consistent for the population parameter σ_b^2. Estimates of the variance of variance component of error terms also decrease with sample size and tend to zero. That is, the estimator, $\hat{\sigma}_e^2$ of the variance component of error terms is consistent for the population parameter σ_e^2.

According to table 2, for both random effects and errors true variance is higher than the estimated variance. The percentage difference for random components is between -84% to -7% and the percentage difference for errors is between -1% to -98% and this does not depend on σ_b^2 nor sample size. For both components, the difference is quite high. That is, it fails to include all the information of the population. Therefore, it can be concluded that the estimators of the variance component of random effects and errors are not sufficient on its own for the population parameters σ_b^2 and σ_e^2 respectively for any value of σ_b^2.

In order to examine the efficiency of REML method, the ratio between MIVQUE lower bound and the variances of variance estimates of random components under REML method is considered. For small σ_b^2 values, the ratios are very close to one. But, for higher values of σ_b^2, the ratio is away from 1 (less than 1). So it can be concluded that REML method is very efficient for small σ_b^2 values and less efficient for higher values of σ_b^2.

Table 3 gives a summary of the results obtained from tables 1 and 2.

Table 3 should come here

4. Discussion

The linear Gaussian mixed model is highly used in the description and analysis of data from varied fields. The great versatility of this models has only relatively recently been generally accessible to users. The key advantage of this mixed model over its fixed effects counterpart is that it can cope with the situation when the data does not satisfy the assumptions of independence and variance homogeneity. The mixed model extends the fixed effects model by including random effects, random coefficients and / or covariance terms in the residual variance matrix. There are several estimation procedures available for estimating the fixed effects components and variance components of the random effects. Many studies provide for parameter estimation based on Maximum Likelihood and Residual Maximum Likelihood with different algorithms available. Literature indicates that Least Squares Estimation method is not appropriate. And for normal data, Iterative Generalized Least Squares and Maximum Likelihood Estimation methods provide the same estimates. Residual Maximum Likelihood Estimation (REML) method is said to be the best procedure for estimating parameters in mixed models.

As the REML estimation procedure is an iterative one the properties of REML estimators cannot be checked analytically. Past simulation studies have only examined the properties of unbiasedness and efficiency of REML estimators and many of these studies concentrate only on either fixed or random effects. This research was aimed at studying the properties of REML estimators for both fixed and random effects. Here, the properties, unbiasedness, consistency, sufficiency, and efficiency of these estimators are checked via a comprehensive simulation study.

The results indicate that the Residual Maximum Likelihood estimation (REML) method holds all the desired properties for fixed effects. However for variance components of random effects and errors it does not hold the property of sufficiency and also though when the ratio of variance of random effects to error variance is small it holds the property of efficiency it is not so efficient when this ratio is large.

References

1.Brown, H. and Prescott, R. (2001), *Applied Mixed Models in Medicine*, New York: John Wiley and Sons, Inc.

2.Hartley, H.O. and Rao, J.N.K. (1967), 'Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model' *Biometrika*, 54 : 93 -108.

3.Harville, D.A. (1977), 'Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems' *Journal of the American Statistical Association*, 72 : 320 -338.

4. Harville, D.A. (1988), 'Mixed-Model Methodology: Theoretical Justifications and Future Directions,' *Proceedings of the Statistical Computing Section*, American Statistical Association, New Orleans, 41 -49.

5.Harville, D.A. (1990), 'BLUP (Best Linear Unbiased Prediction), and Beyond,' in *Advances in Statistical Methods for Genetic Improvement of Livestock*, Springer-Verlag, 239 -276.

6.Kenward, M.G. and Roger, J.H. 1997. 'Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood' .. *Biometrics*, 53(3) : 983-997.

7.Jennrich, R.I. and Schluchter, M.D. (1986), 'Unbalanced Repeated-Measures Models with Structured Covariance Matrices,' *Biometrics*, 42, 805 -820.

8.Laird, N.M. and Ware, J.H. (1982), 'Random-Effects Models for Longitudinal Data' *Biometrics*, 38, 963 -974.

9.Liao, J.G. and Lipsitz, S.R. (2002) 'A type of restricted maximum likelihood estimator for variance components in generalized linear mixed models' *Biometrika* 89, 401-409

10.Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996), *SAS System for Mixed Models,* Cary, NC: SAS Institute Inc.

11. McGilchrist, C.A. 1988. 'Bias of ML and REML estimators in regression models with ARMA errors'. *Journal of Statistical Computation and Simulation*, 32(3) 127 – 136.

12.McLean, R.A. and Sanders, W.L. (1988), 'Approximating Degrees of Freedom for Standard Errors in Mixed Linear Models,' *Proceedings of the Statistical Computing Section*, American Statistical Association, New Orleans, 50 -59.

13.Mood, A.M., Graybill, F.A and Boes, D.C. (1963) '*Introduction to the Theory of Statistics*' McGraw-Hill Series in Probability and Statistics.

14.Patterson, H.D. and Thompson, R. (1971), 'Recovery of Inter-Block Information when Block Sizes are Unequal' *Biometrika*, 58, 545 -554

15.Searle, S. R. (1971), *Linear Models*, New York: John Wiley & Sons, Inc.

16. Swallow, W. H. and Monahan, J. F. 1984. 'Monte Carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components'. *Journal of American Statistical Association and American Society for Quality* 26(1) : 47-57.

17.Swallow, W.H., Searle, S.R. 'Minimum Variance Quadratic Unbiased Estimation (MIVQUE) of Variance Components' *Technometrics* 20(3) : 265-272

18.Rao,C.R. (1971b), 'Minimum Variance Quadratic Unbiased Estimation of Variance omponents' *Journal of Multivariate Analysis* 1, 445-446

19.Robinson, G.K. (1991), 'That BLUP is a Good Thing: The Estimation of Random Effects,' *Statistical Science*, 6, 15 -51.

20.Verbeke, G. and Molenberghs, G., eds. (1997), *Linear Mixed Models in Practice: A SAS-Oriented Approach,* New York: Springer.

21.Verbeke, G. and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data,* New York: Springer.

22. Zhang, Q., Zhang, Y., Liu, Z., Haussmann, H. 1995. 'Comparison of MIVQUE and REML with Monte Carlo simulation' Thesis, College of Animal Science and Technology, Beijing Agricultural University (PMID : 8900839).

Table 1 – Parameter estimate ($\hat{\alpha}$), Bias (B($\hat{\alpha}$)), estimated variance (EV($\hat{\alpha}$)) and percentage difference in variance (P($\hat{\alpha}$)) under REML estimation of Fixed effect(α) for different sample sizes (n)

Case(1) - for σ_b^2 =0.0																
α	n=20				n=60				n=100				n=200			
	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)
1	1.010	0.010	0.00000	0.667	1.004	0.004	0.00001	1.020	0.999	-0.001	0.00001	-2.976	0.996	-0.004	0.000001	1.493
2	2.010	0.010	0.00008	-2.50	2.010	0.010	0.00007	1.449	1.998	-0.002	0.00001	0.000	2.003	0.003	0.000002	-0.641
3	2.999	-0.001	0.00005	0.882	3.006	0.006	0.00004	0.943	2.999	-0.001	0.00002	0.000	2.990	-0.010	0.000004	0.000
4	4.003	0.003	0.00000	0.000	3.997	-0.003	0.00000	-1.613	3.983	-0.017	0.00001	-4.546	3.988	-0.012	0.000000	-2.381
Case(2) - for σ_b^2 =0.1																
α	n=20				n=60				n=100				n=200			
	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)
1	1.009	0.009	4.9E-6	1.04	0.997	-0.003	1.0E-6	0.00	0.997	-0.003	5.6E-6	1.85	1.002	0.002	8.0E-8	0.71
2	2.000	0.000	2.4E-5	3.47	1.999	-0.001	4.2E-6	2.17	2.006	0.006	7.7E-7	-4.26	2.006	0.006	4.3E-7	-2.27
3	2.996	-0.004	1.4E-5	-2.63	2.993	-0.007	1.8E-6	-2.78	2.998	-0.002	1.7E-6	-4.45	3.009	0.009	2.4E-7	3.85
4	4.012	0.012	2.9E-5	-1.67	3.995	-0.005	1.9E-6	-2.50	4.006	0.006	6.4E-6	-0.77	3.999	-0.001	2.0E-8	1.79
Case(3) - for σ_b^2 =0.2																
α	n=20				n=60				n=100				n=200			
	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)
1	1.017	0.017	1.5E-4	0	1.011	0.011	5.8E-5	0.8772	0.987	-0.013	1.5E-5	0	1.022	0.022	1.5E-05	3.5714
2	2.017	0.017	1.5E-4	1.4286	1.993	-0.007	7.7E-7	-2.8571	1.973	-0.027	3.3E-7	3.3333	2.019	0.019	1.5E-4	-3.125
3	3.006	0.006	5.0E-5	-1.087	2.990	-0.010	4.5E-5	2.9412	2.973	-0.027	3.6E-6	0	3.007	0.007	5.0E-5	0
4	3.992	-0.008	2.1E-4	-2.5	3.983	-0.017	6.4E-6	-4.5	3.973	-0.027	9.5E-7	0	4.002	0.002	2.1E-4	0
Case(4) - for σ_b^2 =0.5																
α	n=20				n=60				n=100				n=200			
	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)
1	1.007	0.007	1.4E-5	3.57	1.009	0.009	1.1E-5	0.00	1.001	0.001	5.6E-6	0.00	0.993	-0.007	4.5E-6	0.00
2	2.005	0.005	3.4E-5	0.00	2.002	0.002	3.0E-5	0.00	1.997	-0.003	9.2E-6	-4.00	1.996	-0.004	9.0E-8	-2.94
3	3.006	0.006	1.3E-4	0.00	3.005	0.005	4.0E-5	-4.49	2.999	-0.001	4.6E-6	0.00	2.998	-0.002	1.7E-6	1.52
4	3.996	-0.004	5.4E-5	1.92	4.000	0.000	2.2E-5	0.00	3.995	-0.005	1.7E-5	0.00	3.999	-0.001	1.3E-7	0.00

Table 1 continued ……..

Case(5) - for $\sigma_b^2=1$																
α	n=20				n=60				n=100				n=200			
	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)
1	0.989	-0.011	8.3E-5	-4.630	0.999	-0.001	1.9E-5	-1.667	1.009	0.009	4.9E-6	0.610	1.000	0.000	2.9E-6	0.000
2	1.996	-0.004	8.6E-5	0.000	1.996	-0.004	8.3E-5	0.000	2.006	0.006	6.5E-5	0.862	2.001	0.001	5.9E-5	-0.595
3	3.003	0.003	6.2E-5	-1.111	2.997	-0.003	4.4E-5	-0.794	3.007	0.007	5.7E-6	1.667	2.999	-0.001	8.0E-8	0.000
4	4.001	0.001	2.3E-5	-2.326	3.995	-0.005	8.2E-6	-1.125	4.006	0.006	1.9E-6	-4.500	4.002	0.002	6.2E-7	2.778
Case(6) - for $\sigma_b^2=2$																
α	n=20				n=60				n=100				n=200			
	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)
1	1.017	0.017	3.6E-5	3	1.011	0.011	9.3E-6	2.083	0.987	-0.013	2.5E-6	0	1.022	0.022	2.4E-7	1.667
2	2.017	0.017	1.6E-4	2.778	1.993	-0.007	1.9E-5	0	1.973	-0.027	1.4E-6	0	2.019	0.019	1.7E-7	-4.95
3	3.006	0.006	7.2E-5	-0.833	2.990	-0.010	5.9E-5	0	2.973	-0.027	6.6E-6	0	3.007	0.007	4.5E-6	4.878
4	3.992	-0.008	5.9E-5	0.862	3.983	-0.017	2.4E-5	-2	3.973	-0.027	4.6E-6	-0.660	4.002	0.002	2.8E-6	0.6
Case(7) - for $\sigma_b^2=5$																
α	n=20				n=60				n=100				n=200			
	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)	$\hat{\alpha}$	B($\hat{\alpha}$)	EV($\hat{\alpha}$)	P($\hat{\alpha}$)
1	1.002	0.002	1.0E-3	0	0.999	-0.001	2.0E-4	-2.273	0.996	-0.004	1.7E-4	0.000	1.002	0.002	2.1E-5	0.000
2	1.998	-0.002	6.2E-4	2.381	2.006	0.006	2.2E-4	0.932	1.997	-0.003	1.4E-4	3.846	2.006	0.006	7.0E-7	-0.794
3	3.003	0.003	1.0E-3	0	3.005	0.005	5.1E-4	2.941	3.006	0.006	3.1E-4	1	2.995	-0.005	1.8E-5	-4.5
4	4.003	0.003	7.7E-4	0	3.997	-0.003	1.5E-4	3.333	4.002	0.002	8.9E-5	2.976	3.999	-0.001	1.6E-5	0

Table 2 – Parameter estimates, Bias, estimated variance(EV), percentage difference in variance(PDV) under REML estmation and ratio of variance of REML to MIVQUE lower bound (Ratio) for the random effects for different combinations of sample size (n)

Sample Size	Variance Component	Case (1) : σ^2_b =0.0				
		Estimate	Bias	EV	PDV	Ratio
n=20	$\hat{\sigma}_b^2$	-0.0223	-0.0223	0.0039	-7.0920	0.9291
	$\hat{\sigma}_e^2$	1.0790	0.0790	0.1899	-43.5800	0.9342
n=60	$\hat{\sigma}_b^2$	-0.0029	-0.0029	0.0001	-30.0000	0.8800
	$\hat{\sigma}_e^2$	1.0706	0.0706	0.0718	-18.4600	0.8670
n=100	$\hat{\sigma}_b^2$	0.0114	0.0114	0.0001	-21.4300	0.8755
	$\hat{\sigma}_e^2$	1.0675	0.0675	0.0524	-1.5890	1.0159
n=200	$\hat{\sigma}_b^2$	0.0028	0.0028	0.0000	-66.6700	0.8833
	$\hat{\sigma}_e^2$	1.0031	0.0031	0.0033	-21.4550	1.1146
Sample Size	Variance Components	Case (2) : σ^2_b =0.1				
		Estimate	Bias	EV	PDV	Ratio
n=20	$\hat{\sigma}_b^2$	0.09406	-0.0059	0.00761	-27.593	0.91407
	$\hat{\sigma}_e^2$	1.07412	0.07412	0.18497	-26.976	0.91024
n=60	$\hat{\sigma}_b^2$	0.09399	-0.006	0.00923	-62.214	1.02214
	$\hat{\sigma}_e^2$	1.04556	0.04556	0.08534	-50.141	1.02141
n=100	$\hat{\sigma}_b^2$	0.10641	0.00641	0.00754	-10.664	0.89336
	$\hat{\sigma}_e^2$	1.00513	0.00513	0.0308	-36.1	0.899
n=200	$\hat{\sigma}_b^2$	0.1002	0.0002	0.0057	-12.308	0.87692
	$\hat{\sigma}_e^2$	1.00001	1.00E-05	0.00971	-9.252	0.90748
Sample Size	Variance Components	Case (3) : σ^2_b =0.2				
		Estimate	Bias	EV	PDV	Ratio
n=20	$\hat{\sigma}_b^2$	0.20602	0.00602	0.02799	-32.505	1.00495
	$\hat{\sigma}_e^2$	1.06046	0.06046	0.18472	-45.305	0.87695
n=60	$\hat{\sigma}_b^2$	0.1981	-0.0019	0.02319	-33.439	0.99561
	$\hat{\sigma}_e^2$	1.09861	0.09861	0.0412	-50.175	0.82825
n=100	$\hat{\sigma}_b^2$	0.20384	0.00384	0.01694	-49.372	0.83628
	$\hat{\sigma}_e^2$	1.09605	0.09605	0.02814	-43.776	0.89224
n=200	$\hat{\sigma}_b^2$	0.19978	-0.0002	0.00456	-5.241	0.94759
	$\hat{\sigma}_e^2$	0.98773	-0.0123	0.01922	-29.156	1.03844
Sample Size	Variance Components	Case (4) : σ^2_b =0.5				
		Estimate	Bias	EV	PDV	Ratio
n=20	$\hat{\sigma}_b^2$	0.48495	-0.0151	0.13684	-34.667	0.92333
	$\hat{\sigma}_e^2$	1.03329	0.03329	0.18572	-49.761	0.77239

Table 2 continued………

n=60	$\hat{\sigma}_b^2$	0.51801	0.01801	0.12026	-40.377	0.86623
	$\hat{\sigma}_e^2$	1.0268	0.0268	0.04474	-51.669	0.75331
n=100	$\hat{\sigma}_b^2$	0.48306	-0.0169	0.11479	-44.889	0.82111
	$\hat{\sigma}_e^2$	1.06768	0.06768	0.02633	3.336	1.03336
n=200	$\hat{\sigma}_b^2$	0.50192	0.00192	0.0925	-83.682	0.85318
	$\hat{\sigma}_e^2$	1.00754	0.00754	0.00881	-56.68	1.2668
Sample Size	Variance Components	Case (5) : σ^2_b =1.0				
		Estimate	Bias	EV	PDV	Ratio
n=20	$\hat{\sigma}_b^2$	1.07306	0.07306	0.6643	-12.286	0.87714
	$\hat{\sigma}_e^2$	1.00215	0.00215	0.20143	-52.266	0.87734
n=60	$\hat{\sigma}_b^2$	0.95338	-0.0466	0.57593	-46.471	0.93529
	$\hat{\sigma}_e^2$	1.00316	0.00316	0.05733	-63.561	0.76439
n=100	$\hat{\sigma}_b^2$	0.96683	-0.0332	0.3999	-40.826	0.99174
	$\hat{\sigma}_e^2$	1.00282	0.00282	0.04335	-49.593	0.90407
n=200	$\hat{\sigma}_b^2$	0.98747	-0.0125	0.16649	-55.727	0.84273
	$\hat{\sigma}_e^2$	1.00164	0.00164	0.00207	-89.492	1.09508
Sample Size	Variance Components	Case (6) : σ^2_b =2.0				
		Estimate	Bias	EV	PDV	Ratio
n=20	$\hat{\sigma}_b^2$	2.00435	0.00435	1.18922	-21.639	0.77373
	$\hat{\sigma}_e^2$	0.96643	-0.0336	0.40641	-42.588	0.92412
n=60	$\hat{\sigma}_b^2$	1.91065	-0.0894	1.10964	-32.639	1.02361
	$\hat{\sigma}_e^2$	0.9962	-0.0038	0.18322	-41.818	0.93182
n=100	$\hat{\sigma}_b^2$	1.94213	-0.0579	1.09933	-33.622	1.01378
	$\hat{\sigma}_e^2$	0.97522	-0.0248	0.19376	-33.689	1.01311
n=200	$\hat{\sigma}_b^2$	2.03756	0.03756	1.71431	-23.103	1.11897
	$\hat{\sigma}_e^2$	1.00533	0.00533	0.09758	-97.546	0.79454
Sample Size	Variance Components	Case (7) : σ^2_b =5.0				
		Estimate	Bias	EV	PDV	Ratio
n=20	$\hat{\sigma}_b^2$	5.05548	0.05548	4.31253	-55.708	0.78571
	$\hat{\sigma}_e^2$	0.97856	-0.0214	1.38453	-91.245	0.88754
n=60	$\hat{\sigma}_b^2$	5.06444	0.06444	4.12172	-52.226	0.89774
	$\hat{\sigma}_e^2$	1.04054	0.04054	1.36587	-77.809	0.82191
n=100	$\hat{\sigma}_b^2$	4.93899	-0.061	3.98542	-25.149	0.74851
	$\hat{\sigma}_e^2$	0.98875	-0.0113	1.2175	-48.975	0.84025
n=200	$\hat{\sigma}_b^2$	5.06578	0.06578	3.4337	-13.303	0.81545
	$\hat{\sigma}_e^2$	0.99928	-0.0007	0.87393	-69.576	0.81424

Table 3 – Properties of Fixed and Random effect estimators under REML method of estimation

Effect	Unbiasedness	Consistency	Sufficiency	Efficiency
Fixed Effects	√	√	√	More efficient than IGLS and ML estimators
Random Effects	√	√	X	Efficient when ratio of σ_b^2 to σ_e^2 is small but less efficient when this ratio is large.