Nigerian Journal of Mathematics and Applications
Volume **25**, (2016), 19 − 55. *Printed by Unilorin press*
©*Nig. J. Math. Appl.* *http://www.njmaman.com*

On The Efficiency of Maximum Likelihood and Restricted Maximum Likelihood Estimators For Linear Mixed Effects Models

[1]Yahya W. B., [2]Adeniyi I. A., [2]Olayemi R. M., [2]Dauda K. A., [1]Banjoko A. W., [2]Olorede K. O. and [3]Ezenweke C. P.

Abstract

In this work, the performances of maximum likelihood (ML) and restricted maximum likelihood (REML) estimation of linear mixed effects models were examined. The behaviours and properties of the ML and REML methods of estimating the variance-covariance components in linear mixed effects models were studied via Monte-Carlo experiment. Results revealed that both estimation techniques yielded the same estimates of the fixed effects parameters. However, ML estimates of variance-covariance components, especially for the random effects terms were more biased downwards than the REML counterparts. It was also discovered that the biasness of the ML estimates for the variance-covariance components increase as the number of fixed effects parameters in the model increase.

Received: 02/01/2016, Accepted: 22/02/2016, Revised: 04/03/2016.
2010 *Mathematics Subject Classification.* 62Jxx & 62J12. * Corresponding author.
Key words and phrases. Mixed effects models, Maximum likelihood, Restricted Maximum likelihood, random effects.
[1]Department of Statistics, University of Ilorin, Ilorin, Nigeria.;
E-mail: dr.yah2009@gmail.com, banjoko.aw@unilorin.edu.ng
[2]Department of Statistics & Mathematical Sciences, Kwara State University, Malete, Nigeria.
E-mail: kazeem.dauda@kwasu.edu.ng;isaac.adeniyi@kwasu.edu.ng;kabir.olorede@kwasu.edu.ng
[3]Department of Mathematics, Federal University Lokoja, Nigeria.
E-mail: chinenyeezenweke@yahoo.com

1. Introduction

Statistical methods for analyzing cross-sectional data where only one measurement is made for a variable of interest for each individual are well developed, such as classical linear regression models and generalized linear models. An important assumption for cross-sectional data is that observations are independent of each other. Therefore, statistical methods for analysing cross-sectional data cannot directly be used for analysing clustered (e.g. longitudinal) data or correlated data (Pinheiro and Bates, 2000). Some of the typical characteristics of clustered dataset are that it has more than one measurement per cluster and are often unbalanced; there may be significant variation among measurements within each cluster and among measurements between different clusters; the observed data are often complex or incomplete due to the fact that there may be missing data, measurement errors, censoring, and outliers. Because of these characteristics of clustered data, statistical models and methods for analyzing such (longitudinal) data can be challenging. Mixed models were developed to handle clustered data that are characterized with repeated measurements in which data fall into different clusters. Data within the same cluster may be correlated, but data between different clusters are often assumed to be independent (Wu, 2010). In these models random effects are introduced to incorporate the between-individual variation and within-individual correlation in the data. The parameters in a mixed effects model can be classified into two types: fixed effects that are associated with the average effect of predictors on the response, and variance-covariance components associated with the covariance structure of the random effects and of the error term (Wang and Heckman, 2009). Several estimation methods have been proposed for estimating mixed effects models but the maximum likelihood (ML) and restricted maximum likelihood (REML) are generally adopted estimation techniques for linear mixed effects models (Harville, 1977; Longford, 1993). The ML and the REML estimation techniques are the generally acceptable methods of fitting linear mixed effects (LME) models with REML being less biased than ML (Corbeil and Searle, 1976; Haville, 1977). The theoretical reasons why the ML estimates of the variance components are more biased than the REML estimates were provided separately in the above works. The ML estimates are biased downward because the degrees of freedom lost in the estimation of the fixed effects parameters are not incorporated in the estimation of the variance components. Haville (1977) also reported that the mean square error (MSE) of ML estimates may be higher than that of REML estimates. However, no empirical results were used to substantiate these findings.

In the present study, the performances of both the ML and REML estimation methods were investigated on classical LME models with a view to comparing their efficiency on the parameters of the models using Monte Carlo experiments. The behaviour of the two estimators under different numbers of fixed and random effects parameters in the model at varying sample sizes was examined. Knowledge of this would provide a reasonable guide on when to be more concern on the efficiency of the model given the inherent number of fixed and random effects in the model in real life cases. In all cases, the efficiency of the models was assessed using the absolute bias and the mean square errors of the parameters estimates. In section 2, the general concept of Linear Mixed Effects models, the ML and REML estimation techniques are summarized. Section 3 presents an overview of the Monte-Carlo experiment and the results obtained. Detailed discussion of the results and conclusion are provided in Section 5.

2. MATERIAL AND METHODS

2.1. **THE LINEAR MIXED EFFECT MODELS.** A general LME model can be written as

$$y_{ij} = X_i\beta + Z_ib_i + e_{ij}, i = 1, 2, ..., n; j = 1, 2, ..., m \tag{1}$$

$b_i \sim N(0, D)$, $e_{ij} \sim N(0, R_i)$, where y_{ij} is the j^{th} observation for the i^{th} cluster; β is a $Px1$ vector of fixed effects; b_i is a $qx1$ vector of a random effects; X_i is n_ixq design matrix which may also contain covariates. The e_{ij} in model (1) represent the random errors of the measurements within cluster i, D is a qxqcovariance matrix of the random effects, and R_i is a n_ixn_i covariance matrix of the within-cluster errors. We often assume that $R_i = \sigma_e^2 I$ for simplicity, where I is the n_ixn_i identify matrix, i.e the within-cluster measurement are assumed to be independent with constant variance. The value of σ_e^2 represent the magnitude of the within-cluster variation, and the values of the diagonal elements of D represent the magnitude of the between-cluster variation. A distinguishing characteristic of the MLE model (1) is that it is linear in both the mean parameters β and the random effects.

Likelihood methods are the classical inferential tools for mixed effects models. The ML and REML are the most common estimation methods used for linear mixed effects models. The estimates of the unknown parameters in LME model based on the ML and REML estimation techniques can be obtained using an iterative algorithm such as an expectation-maximization (EM) algorithm (Wu, 1983; McLachlan and Krishnan, 1997) or a Newton-Raphson method (Laird and Ware 1982; Lindstrom and Bates 1988; Verbeke and Molenberghs 2001; Harville 1977). Computational details of all these can be find in Pinheiro and Bates (2000).

However, following Pinheiro and Bates (2000), the marginal distribution of the response y is given as $y \sim N(X\beta, ZDZ^T + \sigma_e I)$.The likelihood fucntion of the observed data $y = (y_1, y_2, ..., y_n)$ is given as,

$$L(\theta \mid y) = \prod_{i=1}^{n} f(y_i \mid \beta, \sigma_e^2 I, D) \tag{2}$$

$\rightarrow$

$$L(\theta \mid y) = \prod_{i=1}^{n} f(y_i \mid \beta, \sigma_e^2 I, b_i) f(b, D) db_i \tag{3}$$

The variance of teh error term, σ_e^2 can be factored from variance covariance matrix of the random effect, that is $D = \sigma_e^2 D^s$, where,D^s is called the scaled variance-covariance matrix of the random effects. The log likelihood function of y is then given by
(4)

$$L(\beta, \sigma^2, D^s, \mid y) = \frac{1}{2}\left[n \log(2\pi\sigma_e^2) + log(|D|) + \frac{1}{\sigma^2}(y - X\beta)^T D^{-1}(y - X\beta)\right]$$

where
$D = I + ZD_A^S Z^T$

2.1.1. *Maximum Likelihood (ML) Estimation.* For fixed D^s, the values of β and σ^2 that maximize the log likelihood function of y is given by

$$\hat{\beta}(D^s) = \left[X^T(D)^{-1X}\right]^{-1} X^T(D)^{-1}y \tag{5}$$

$$\sigma_e^2(D^s) = \frac{1}{n}\left[y - X\hat{\beta}(D^s)\right]^T (D)^{-1}\left[y - \hat{\beta}(D^s)\right] \tag{6}$$

2.1.2. *Restricted Maximum Likelihood (REML) Estimation.* The estimates of D and σ_e^2 tend to be underestimated by the ML estimation procedure (Harville, 1977). Therefore, the REML (Harville, 1974) may be preferred for these quantities. Consider the QR decomposition (Thisted, 1988) of X of the form;
$X = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$
where R_1 is the upper triangular. The REML estimates can be obtained from the likelihood of $y^* = Q_2^T y$. It can be shown that $y^* \dim N(0, \sum)$, where $\sum^* = \sigma^2(I + Q_2^T Z D_A^s Z^T Q_2)$. Letting $Z^* = Q_2^T Z$ and $n^* = n - p_0$. The corresponding restricted likelihood can be written as

$$\ell_R(\beta, \sigma^2, D^2 \mid y) = \frac{-1}{2}\left[n^* \log(2\pi\sigma_e^2) + log(|D^*|) + \frac{1}{\sigma^2} y^{*T}(D^*)^{-1}y^*\right] \tag{7}$$

Where
$D^* = I + Z^* D_A^S Z^{*T}$.
For fixed D^S, the value of σ^2 that maximizes the log likelihood is given as:

$$\hat{\sigma}_R^2(D^S) = (\frac{1}{n^*}) y^{*T}(I + Z^* D_A^S Z^{*T}) y^* \tag{8}$$

The restricted likelihood (7) does not depend on β and hence no fixed effects REML estimates are available. However, replacing D^S by its corresponding

REML estimate can be used to provide estimates for the fixed effects in restricted maximum likelihood estimation. Details of this procedure were provided in Hartley and Rao (1967), Laird and Ware (1982), Lindstrom and Bates (1988) and Wolfinger, Tobias and Sall, (1991) among others.

2.2. **SIMULATION SCHEME.** In this section, the scheme adopted for the Monte Carlo experiments were provided and results from the analysis presented. Three different models with varying number of fixed effects parameters were considered. The actual values of the fixed effects and values of the variance components were arbitrarily chosen.

Model 1: This model has one random effect and three fixed effect parameters.

$y_{ij} = (\beta_0 + b_i) + \beta_1 X_1 + \beta_2 X_2 + \epsilon_{ij}, i = 1, 2, ..., n; j = 1, 2, ..., 4$

where, y_{ij} is the j^{th} observation for the i^{th} cluster; $b_i \sim N(0, \sigma_b^2)$ and $\epsilon_{ij} \sim N(0, \sigma_e^2)$, with b_i independent of ϵ_{ij}. The values of the model parameters were set as follows: $\beta_0 = 17.7$,$\beta_1 = 0.7$,$\beta_2 = -3.3$, $\sigma_b^2 = 5.3$, $\sigma_e^2 = 4$. Each cluster n was varied from $3 to 500$.

Model 2: This model has one random effect and four fixed effect parameters.

$y_{ij} = (\beta_0 + b_i) + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon_{ij}, i = 1, 2, ..., n; j = 1, 2, ..., 4$

The difference between this model and model 1 is the extra fixed effect parameter ($\beta_3 = 5$). The settings and values of all other parameters in model 2 are as specified for model 1.

Model 3: This model has one random effect and eight fixed effect parameters.

$y_{ij} = (\beta_0 + b_i) + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \epsilon_{ij}, i = 1, 2, ..., n; j = 1, 2, ..., 4$

This model differs from model 1 and model 2 in that it has eight fixed effects parameters. The values of the additional fixed effects parameters are set as follows: $\beta_4 = 5.1$, $\beta_5 = -5.8$, $\beta_6 = 2.1$, $\beta_7 = -3.5$, $\sigma_b^2 = 5.3$, $\sigma_e^2 = 4$. The setting and values of all other parameters in model 4 are as specified for models 1 and 2.

In all the three models settings above, the numbers of clusters n considered are 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400 and 500 making a total of 22 clusters with 4 observations per cluster. This implies that there are 4n observations per simulation. Also, each model simulated at a particular cluster (sample) size was replicated 1000 times. This implied that a total of 1000 datasets were simulated for each model.

The estimates of the parameters of the models were obtained using the ML and REML procedures. The expected values of the parameter estimates for each model at a particular sample size were obtained by obtaining the average of all the 1000 estimated parameters. The data simulation and all numerical computations were carried out using R statistical package. The R library package "$lme4$" was invoked to obtain the ML and REML techniques.

The relative efficiency of the two estimators (MLE and REML) were examined using absolute bias (AB) and mean square error (MSE) of the estimates as well as Mean square prediction error (MSPE) of the fitted models. Relevant graphical illustrations were provided to give more credence to the numerical results.

3. ANALYSIS AND RESULTS

3.1. **Results.** Relevant results obtained from the simulation studies are provided in this section under the three models specified in Section 2.

3.1.1. *Monte-Carlo Results for Model 1.* In Model 1, there is only one random effect and three fixed effect parameters. The results of ML and REML estimations for the three fixed effect parameters β_0,β_1 and β_2 in the model are presented in Table 1. The results reported in Table 1 are the average estimates of each parameters by MLE and REML methods at each cluster size over 1000 replications. The graphs of absolute bias of the estimated parameters β_0, β_1 and β_2 by MLE and REML methods bases on the results in Table 1 are plotted as presented by Figures 1, 2 and 3 respectively.
One unique feature observed in the results in Table 1 is that the estimates of the fixed effect parameters provided by both the MLE and REML methods were the same for each parameter at each cluster size n.

Table 1: Results of ML and REML estimations of the fixed effects parameters β_0,β_1 and β_2 in Model 1. The average estimates of the parameters using ML and REML estimators are reported in the table for different number of cluster (n)

Number of clusters (n)	β_0 (17.7)		β_1 (0.7)		β_2 (-3.3)	
	ML	REML	ML	REML	ML	REML
5	17.5936	17.5936	0.7058	0.7058	-3.2315	-3.2315
6	17.7912	17.7912	0.6974	0.6974	-3.3253	-3.3253
7	17.6464	17.6464	0.6996	0.6996	-3.2488	-3.2488
8	17.7382	17.7382	0.6984	0.6984	-3.3521	-3.3521
9	17.7039	17.7039	0.7002	0.7002	-3.2865	-3.2865
10	17.7327	17.7327	0.6941	0.6941	-3.2144	-3.2144
20	17.6896	17.6896	0.6994	0.6994	-3.2692	-3.2692
30	17.6801	17.6801	0.7014	0.7014	-3.2800	-3.2800
40	17.7215	17.7215	0.6993	0.6993	-3.2997	-3.2997
50	17.6759	17.6759	0.7003	0.7003	-3.2925	-3.2925
60	17.6844	17.6844	0.6986	0.6986	-3.2734	-3.2734
70	17.7216	17.7216	0.7001	0.7001	-3.3113	-3.3113
80	17.7060	17.7060	0.6993	0.6993	-3.2767	-3.2767
90	17.7014	17.7014	0.7001	0.7001	-3.3081	-3.3081
100	17.6688	17.6688	0.7023	0.7023	-3.3138	-3.3138
200	17.6874	17.6874	0.7005	0.7005	-3.2908	-3.2908
300	17.7077	17.7077	0.6995	0.6995	-3.2980	-3.2980
400	17.7052	17.7052	0.6998	0.6998	-3.3091	-3.3091
500	17.7017	17.7017	0.6998	0.6998	-3.2896	-3.2896

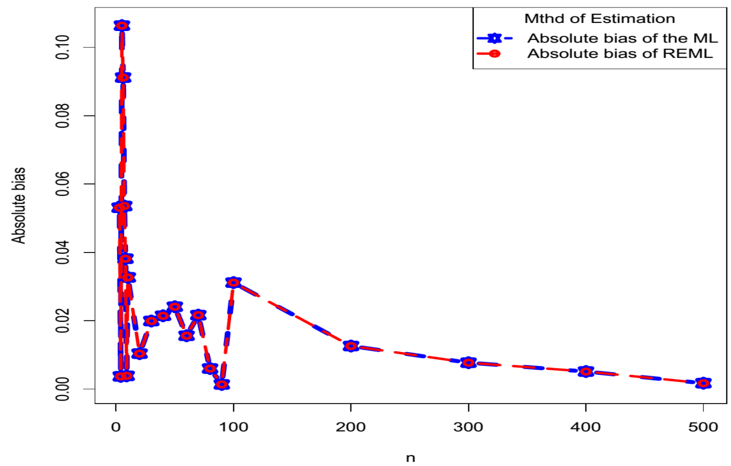


Figure 1: Plots of the absolute bias of ML and REML estimates of β_0 against the cluster sizes n = 5 to n = 500 for Model 1.

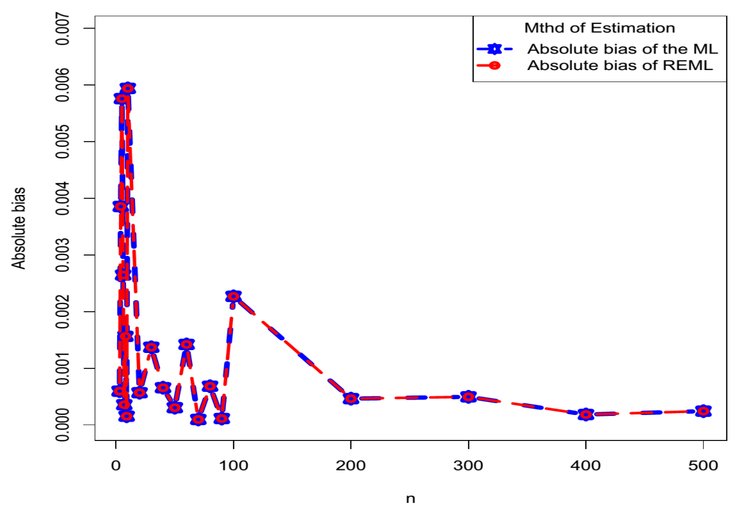


Figure 1: Plots of the absolute bias of ML and REML estimates of β_0 against the cluster sizes n = 5 to n = 500 for Model 1.

Table 2 presents the results of the average estimates of the variance of the random effect parameter σ_b^2 as estimated by ML and REML methods over 1000 replications. The absolute bias and the MSE of the parameter estimates under

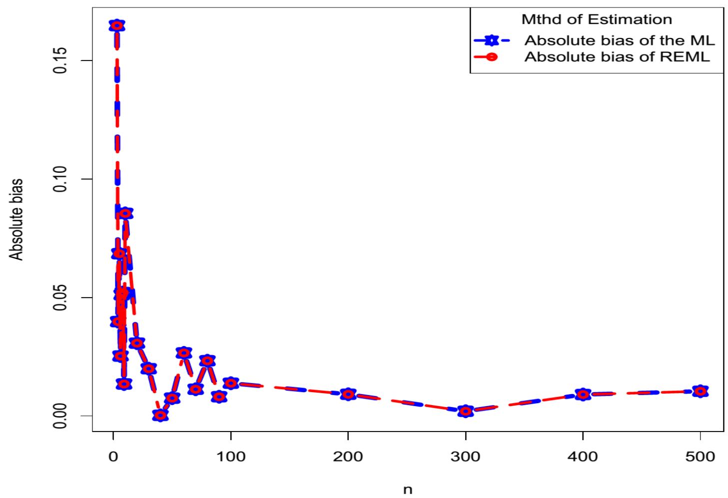
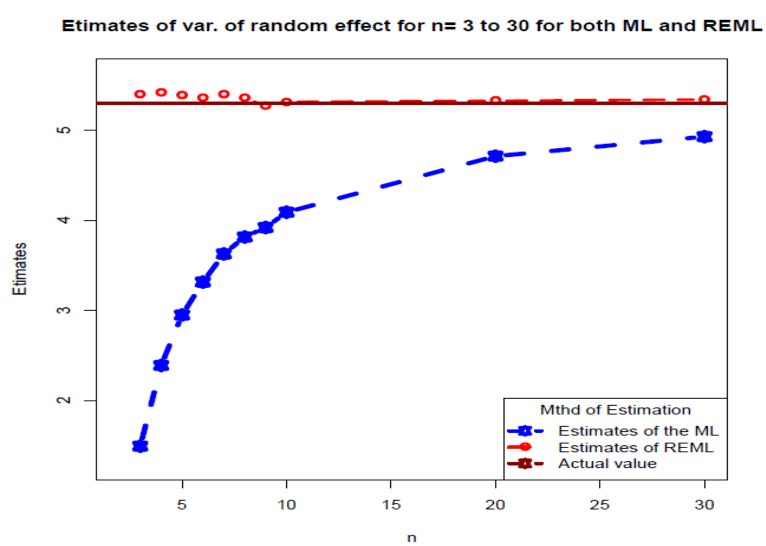


Figure 3: Plots of the absolute bias of ML and REML estimates ofβ_2 against the cluster sizes n = 5 to n = 500 for Model 1.

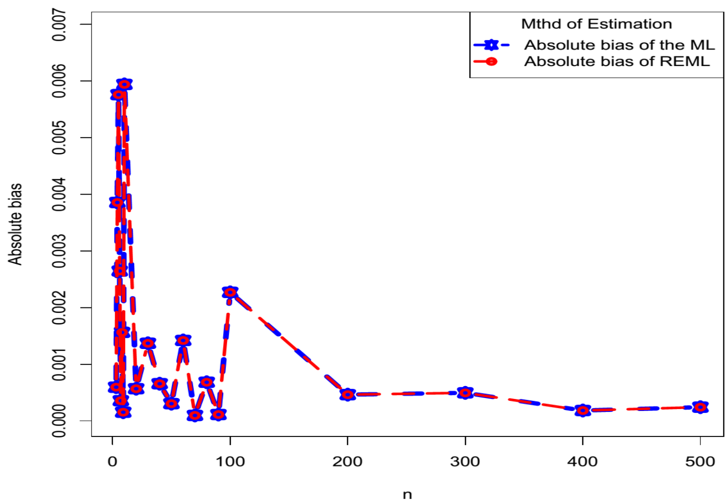
the two estimators are equally provided in the table.
The results in Table 2 reveal that the estimates of σ_b^2 by the ML method are more biased than those of the REML method especially at small sample (cluster) sizes. However, as the sample size increases, the performances of the two methods become apparently similar. This pattern of performance was made clearer by the plot of the estimated σ_b^2 by the ML and REML methods against the cluster size (n) up to n = 30 (to ensure clarity) as presented by the graphs in Figure 4. The graphs of the estimated bias and MSE of this parameter estimates over this small cluster size are also provided as Figure 5 and Figure 6 respectively. Interestingly, the graph in Figure 6 revealed that the MSEs of the ML estimator are relatively lower than those of the REML estimator, especially at lower sample sizes.
Without loss of generality, the plots of the estimated σ_b^2 , its respective bias and MSE over all the 22 clusters employed here are provided by the graphs in Figures A.1, A.2 and A.3 in Appendix A.

Table 2: Results of ML and REML estimations of the variance component σ_b^2 of the random effects parameter in Model 1. The average estimates of σ_b^2 using ML and REML estimators, their respective estimated bias and MSE over 1000 iterations are reported in the table for different number of cluster (n).

Number of clusters (n)	σ_b^2 (5.3)		Absolute Bias		MSE	
	ML	REML	ML	REML	ML	REML
3	1.5268	5.4436	3.7732	0.1436	22.5107	81.3545
4	2.2757	5.1867	3.0244	0.1133	17.6895	35.5712
5	3.1090	5.6594	2.1910	0.3594	15.4717	30.3294
6	3.3602	5.4256	1.9398	0.1256	13.4559	21.9264
7	3.5038	5.2204	1.7962	0.0796	11.2305	15.7791
8	3.7175	5.2244	1.5826	0.0756	10.1687	13.6028
9	3.9269	5.2844	1.3731	0.0156	8.3330	10.6473
10	4.0614	5.2801	1.2386	0.0199	7.8033	9.7759
20	4.6570	5.2681	0.6430	0.0319	3.7828	4.1574
30	4.8655	5.2724	0.4345	0.0276	2.5951	2.7591
40	5.0136	5.3216	0.2864	0.0216	1.9312	2.0471
50	4.9971	5.2399	0.3029	0.0601	1.7681	1.8220
60	5.0689	5.2724	0.2311	0.0276	1.3014	1.3358
70	5.0973	5.2717	0.2027	0.0283	1.1266	1.1507
80	5.1306	5.2836	0.1694	0.0164	0.9958	1.0171
90	5.1580	5.2943	0.1421	0.0057	0.9051	0.9255
100	5.1956	5.3186	0.1044	0.0186	0.7638	0.7840
200	5.2542	5.3157	0.0458	0.0157	0.4121	0.4184
300	5.2695	5.3105	0.0305	0.0105	0.2783	0.2812
400	5.2892	5.3199	0.0108	0.0199	0.2250	0.2275
500	5.2697	5.2942	0.0303	0.0058	0.1454	0.1456

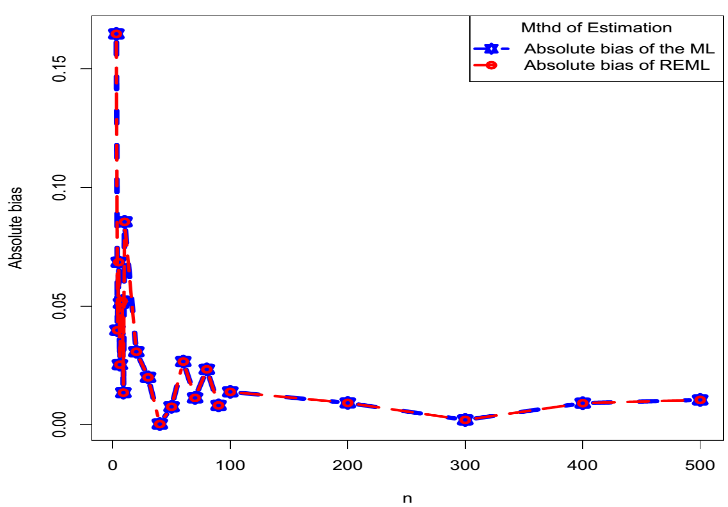


[htp]
Figure 4: Plots of the estimated σ_b^2 for Model 1 using ML and REML methods against the cluster sizes n up to n = 30.



[htp]

Figure 5: Plots of the absolute bias of ML and REML estimates of σ_b^2 for Model 1 against the cluster sizes n up to n = 30.



[htp]

Figure 6: Plots of the MSE of ML and REML estimates of σ_b^2 for Model 1 against the cluster sizes n up to n = 30.

Also, the results of the MLE and REML estimations of the variance of the error term in Model 1 σ_b^2 are provided in Table 3. Graphs in Figures 7, 8 and 9 are the plots of the estimated values of σ_b^2 , the respective bias and MSE over the cluster

sizes (n) up to n = 30. Also, for large values of n, the plots of the estimated σ_b^2, its respective bias and MSE over all the 22 clusters employed are provided by the graphs in Figures A.4, A.5 and A.6 in Appendix A.
It can be observed from the results in Table 3 and Figures 7 to 9 that the estimates of σ_b^2 provided by REML method are closer to the true value of 4 used for simulating the model. Thus, the ML estimator appeared more biased than the REML estimator at estimating this variance parameter. However, the estimates provided by the two methods become closer as the sample size increases.
Finally, the efficiencies of both the ML and REML methods at estimating Model 1 were accessed by computing the MSPE of the fitted model. This was done to determine the prediction strength of the fitted models based on the ML and REML estimators. The plot of the computed average MSPE against the sample (cluster) sizes averaged over 1000 iterations is provided by the graph in Figure 10. The graphs of MSPE of the two estimators over all 22 cluster samples are presented as Fig A.7 in Appendix A.

Table 3: Results of ML and REML estimations of the variance component σ_b^2 of the random error term Model 1. The average estimates of σ_b^2 using ML and REML estimators, their respective average estimated bias and MSE over 1000 iterations are reported in the table for different number of clusters (n).

Number of clusters (n)	**σ_e^2 (4)**		**Absolute Bias**		**MSE**	
	ML	**REML**	**ML**	**REML**	**ML**	**REML**
3	3.1544	3.8222	0.8456	0.1778	3.1156	3.5997
4	3.4940	3.9093	0.5060	0.0907	2.3529	2.6643
5	3.5786	3.8729	0.4214	0.1271	2.0896	2.2819
6	3.6931	3.9272	0.3069	0.0728	1.6612	1.7951
7	3.7496	3.9475	0.2504	0.0525	1.5231	1.6263
8	3.7678	3.9359	0.2322	0.0641	1.2922	1.3597
9	3.8508	4.0010	0.1492	0.0010	1.2299	1.3029
10	3.8248	3.9591	0.1752	0.0409	1.0587	1.1064
20	3.9826	4.0502	0.0174	0.0502	0.5059	0.5256
30	3.9449	3.9892	0.0551	0.0108	0.3434	0.3482
40	3.9860	4.0195	0.0140	0.0195	0.2623	0.2669
50	3.9585	3.9851	0.0415	0.0149	0.2219	0.2234
60	3.9795	4.0017	0.0205	0.0017	0.1726	0.1741
70	3.9648	3.9838	0.0352	0.0162	0.1486	0.1491
80	3.9988	4.0156	0.0012	0.0156	0.1381	0.1395
90	4.0141	4.0291	0.0141	0.0291	0.1142	0.1157
100	3.9814	3.9947	0.0186	0.0053	0.1113	0.1117
200	3.9942	4.0009	0.0058	0.0009	0.0547	0.0548
300	3.9883	3.9928	0.0117	0.0072	0.0334	0.0334
400	3.9947	3.9981	0.0053	0.0019	0.0275	0.0275
500	4.0026	4.0052	0.0026	0.0052	0.0219	0.0220

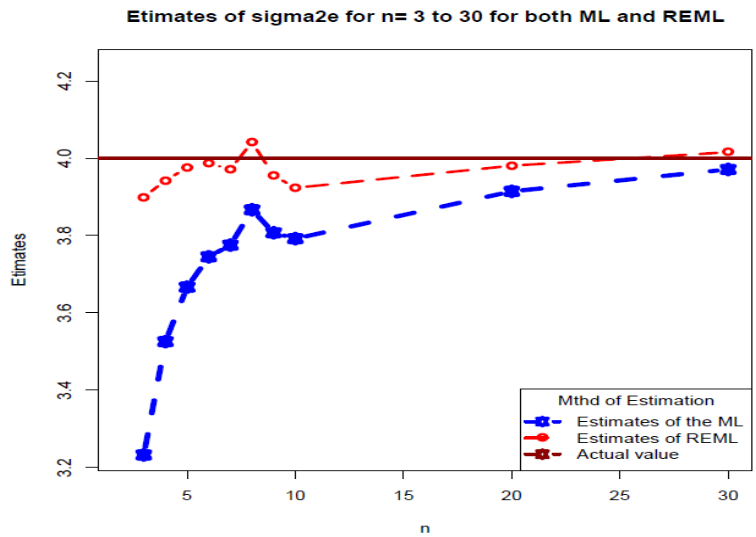


Figure 7: Plots of the estimated σ_b^2 for Model 1 using ML and REML methods against the cluster sizes n up to n = 30.

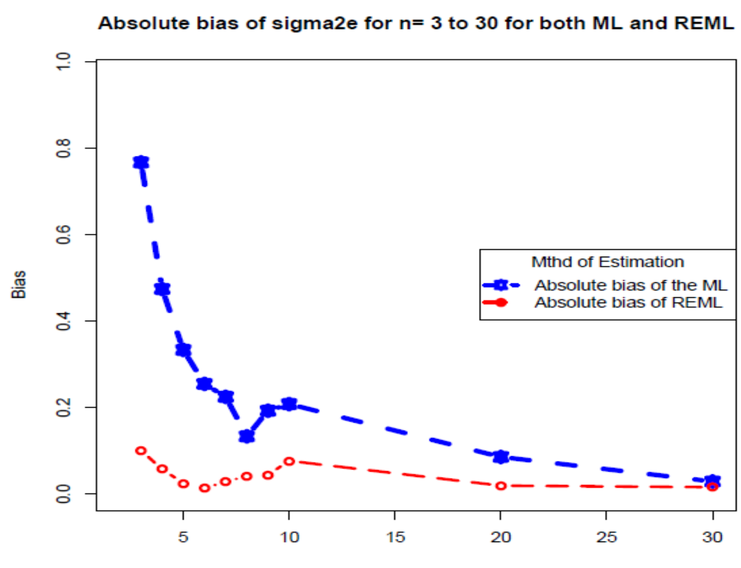


Figure 8: Plots of the absolute bias of ML and REML estimates of σ_b^2 for Model 1 against the cluster sizes n up to n = 30.

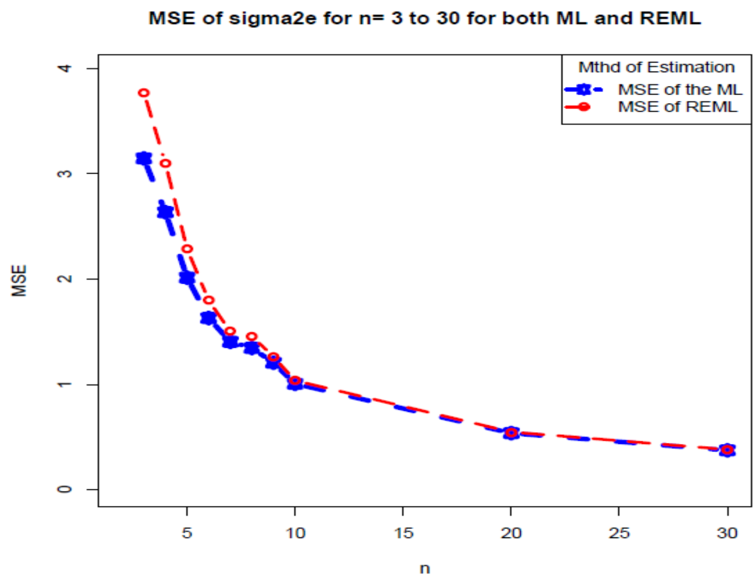
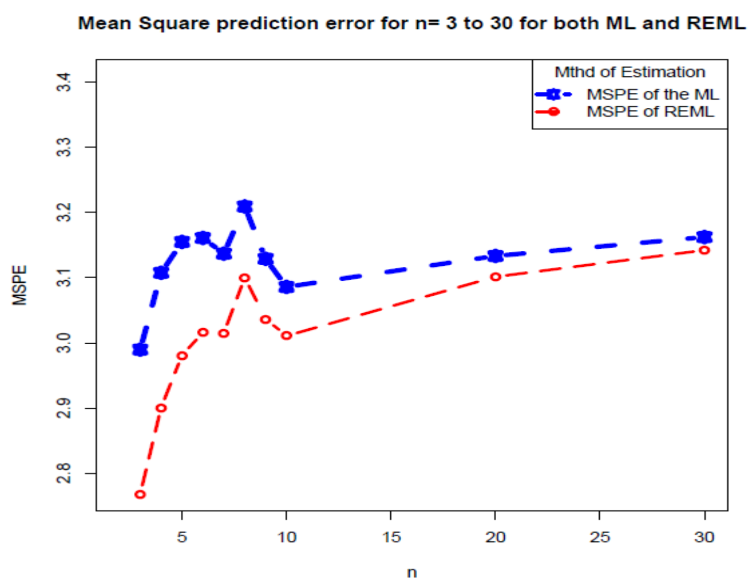


Figure 9: Plots of the MSE of ML and REML estimates of σ_b^2 for Model 1 against the cluster sizes n up to n = 30.



[htp]

Figure 10: Plots of the MSPE against the cluster sizes n up to n = 30 based on ML and REML estimators for Model 1.

3.2. **Monte-Carlo Results for Model 2.** Model 2 has one random effect and four fixed effects parameters. Additional fixed effects parameter $\beta_3 = 5)$ was included in Model 1 to have Model 2 while the remaining terms in the model were

left unaltered. The behaviours of both the ML and REML estimators on fixed effects parameters were the same as those of the first model, Model 1. Hence, the results of fixed effects parameters estimates for Model 2 using the ML and REML methods are not reported here again. Henceforth, further discussions shall be on estimation results for the variance of the random effects components σ_b^2 and σ_e^2 in the model.
Summary of the results of estimations of σ_b^2 and σ_e^2 in Model 2 by both the ML and REML methods are provided in Tables 4 and 5. The two tables presented the estimated average values of σ_b^2 and σ_e^2 by ML and REML estimators, their respective absolute biases and MSEs over 1000 iterations. The results from the two tables equally showed that the REML estimator is relatively more efficient that the ML estimator using the computed absolute biases and MSEs of these estimators, especially at small sample sizes. The performances of the ML estimator equally improved as the sample size increases.

Table 4: Results of ML and REML estimations of the variance component σ_b^2 of the random effects parameter in Model 2. The average estimates of σ_b^2 using ML and REML estimators, their respective estimated bias and MSE over 1000 iterations are reported in the table for different number of cluster (n).

Number of clusters (n)	σ_b^2 (5.3)		Absolute Bias		MSE	
	ML	REML	ML	REML	ML	REML
3	0.0000	5.0307	5.3000	0.2693	28.0900	7.6507
4	1.0272	5.5149	4.2728	0.2149	21.9843	70.4537
5	1.6692	5.1599	3.6308	0.1401	18.9950	38.9301
6	2.3239	5.4222	2.9761	0.1222	15.0860	25.8774
7	2.7663	5.4642	2.5337	0.1642	12.6687	19.4695
8	3.0014	5.3024	2.2986	0.0024	11.5745	16.2832
9	3.2111	5.2472	2.0889	0.0528	10.5457	13.9662
10	3.4693	5.3305	1.8307	0.0305	9.3093	12.1613
20	4.4029	5.3376	0.8971	0.0376	4.4523	5.0402
30	4.6752	5.2942	0.6248	0.0059	2.8850	3.0754
40	4.8119	5.2739	0.4881	0.0261	2.1215	2.1995
50	4.9568	5.3298	0.3432	0.0298	1.4328	1.4881
60	5.0323	5.3441	0.2677	0.0441	1.3137	1.3777
70	5.0318	5.2968	0.2682	0.0032	1.1429	1.1685
80	5.1386	5.3736	0.1614	0.0736	1.0069	1.0636
90	5.0874	5.2935	0.2126	0.0065	0.8642	0.8758
100	5.0650	5.2491	0.2350	0.0509	0.8264	0.8218
200	5.2146	5.3076	0.0854	0.0076	0.4000	0.4048
300	5.2125	5.2741	0.0875	0.0259	0.2631	0.2613
400	5.2464	5.2928	0.0536	0.0072	0.1956	0.1957
500	5.2627	5.2999	0.0373	0.0001	0.1654	0.1660

The plots of estimated values of σ_b^2 by the ML and REML methods against the cluster size (n) up to n = 30 (for clarity) is provided by the graph in Figure 11. The graphs of the estimated σ_b^2 by both methods over the entire cluster numbers n (up to n = 500) is presented as Figure B.1 in Appendix B.

Table 5: Results of ML and REML estimations of the variance component σ_b^2 of the random error term Model 2. The average estimates of σ_b^2 using ML and REML estimators, their respective average estimated bias and MSE over 1000 iterations are reported in the table for different number of clusters (n).

Number of clusters (n)	σ_e^2 (4)		Absolute Bias		MSE	
	ML	REML	ML	REML	ML	REML
3	2.7267	4.0900	1.2733	0.0900	3.5082	4.2534
4	3.3043	3.9371	0.6957	0.0629	2.5687	2.9515
5	3.5912	3.9955	0.4088	0.0045	1.7924	2.0456
6	3.7408	4.0243	0.2593	0.0243	1.6093	1.8186
7	3.8074	4.0265	0.1926	0.0265	1.6048	1.7856
8	3.7793	3.9619	0.2207	0.0381	1.3506	1.4602
9	3.8297	3.9855	0.1703	0.0145	1.1917	1.2612
10	3.8667	4.0052	0.1333	0.0052	0.9788	1.0351
20	3.9553	4.0223	0.0447	0.0223	0.5570	0.5744
30	3.9823	4.0270	0.0177	0.0270	0.3634	0.3720
40	3.9539	3.9871	0.0461	0.0129	0.2908	0.2937
50	3.9628	3.9894	0.0372	0.0106	0.2034	0.2048
60	3.9887	4.0110	0.0113	0.0110	0.1809	0.1829
70	3.9676	3.9866	0.0324	0.0134	0.1429	0.1434
80	4.0050	4.0217	0.0050	0.0217	0.1301	0.1316
90	3.9898	4.0046	0.0102	0.0046	0.1195	0.1203
100	3.9790	3.9923	0.0210	0.0077	0.1161	0.1165
200	3.9936	4.0002	0.0064	0.0002	0.0528	0.0530
300	3.9934	3.9978	0.0066	0.0022	0.0362	0.0362
400	3.9933	3.9967	0.0067	0.0033	0.0255	0.0255
500	3.9916	3.9942	0.0085	0.0058	0.0207	0.0207

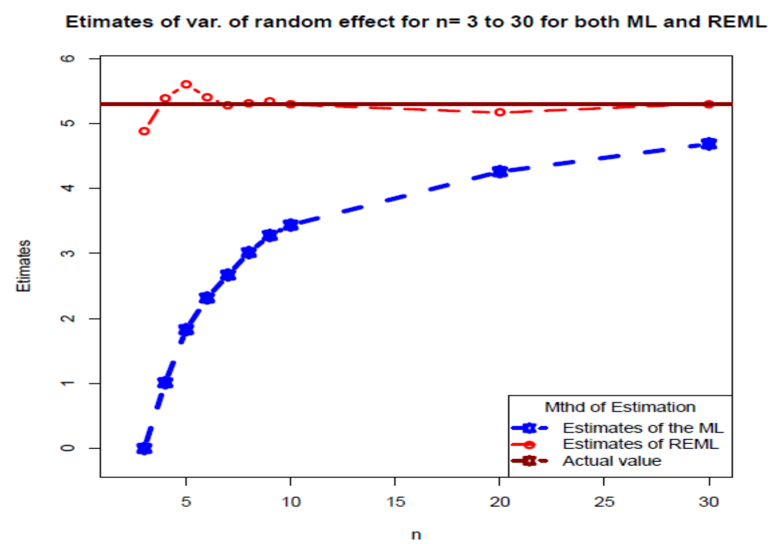


Figure 11: Plots of the estimated σ_b^2 for Model 2 using ML and REML methods against the cluster sizes n up to n = 30.

The graphs of the estimated bias and MSE of σ_b^2 by ML and REML methods over the entire cluster sizes are also provided as Figure 12 and Figure 13 respectively.

From the graphs in Figures 11 and 13, it can be observed that the REML estimator with relatively smaller absolute bias is more efficient than the ML estimator for σ_b^2. Although, the graph of the MSE in Figure 13 indicated a slight improvement of ML estimator over the REML method, especially at smaller sample sizes. Nonetheless, the performances of the two methods are apparently similar at higher sample sizes.

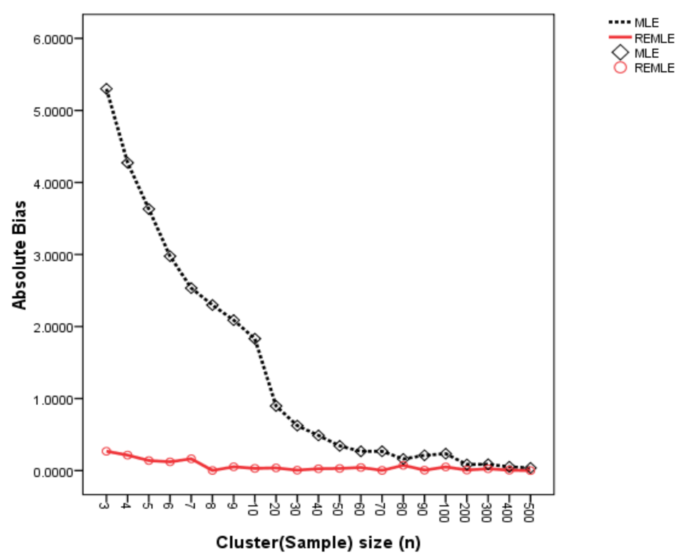


Figure 12: Plots of the absolute bias of ML and REML estimates of σ_b^2 for Model 2 against all the cluster (sample) sizes n (up to n = 500).

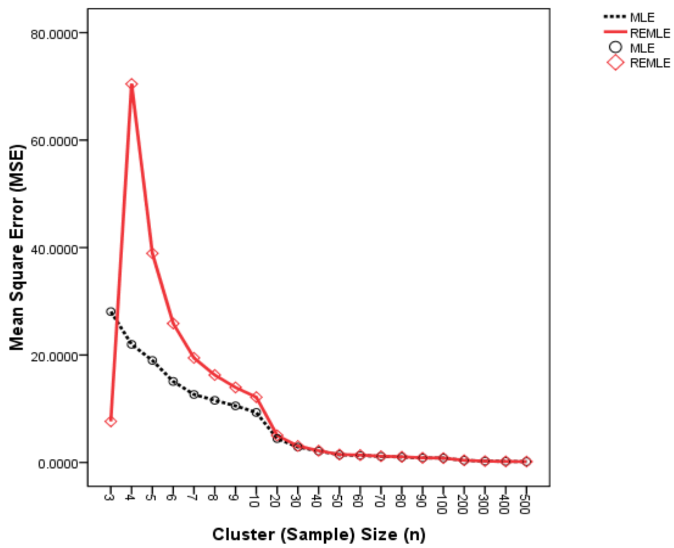


Figure 13: Plots of the MSE of ML and REML estimates of σ_b^2 for Model 2 against all the cluster (sample) sizes n (up to n = 500).

Similarly, the graphs of the estimates of σ_e^2 by ML and REML estimators for Model 2 over the cluster size n up to n = 30 is given by Figure 14. The graphs of the absolute biases and MSEs of the estimates of σ_e^2 of Model 2 by the two estimators are also provided by Figures 15 and 16 respectively. Also, the plots of MSPEs of the ML and REML estimators for Model 2 are also provided as Figure 17.

The behaviours of ML and REML estimators on σ_e^2 for Model 2 are quite similar to their results for Model 1 in which the biases of the REML estimators are relatively smaller than those of the ML estimation method. Also, like in the case of the first model, the MSEs of the ML estimates are lower than those of the REML as observed in Table 5 and 6. This can also be observed in the Fig. 13 and 16.

The results in Table 5 and the graphs of the results given by Figures 14, 15 and 16 showed that the behaviour of the estimates for σ_e^2 is very similar to those of the first model. Once again, the MSPEs yielded by models fitted by REML are lower than those fitted by ML as shown by the graphs in Figure 17 for cluster sizes n up to n = 30. The graph of MSPEs over all the sample sizes for Model 2 is presented as Figure B.3 in Appendix B.

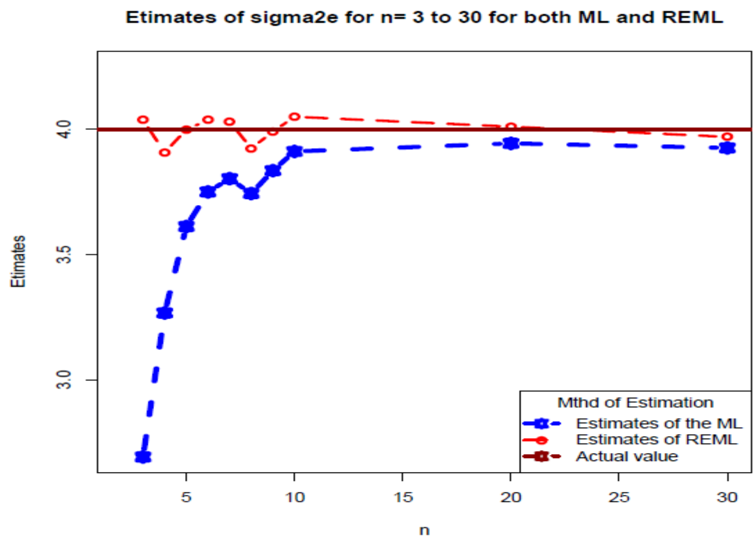


Figure 14: Plots of the estimated σ_e^2 for Model 2 using ML and REML methods against the cluster sizes n up to n = 30.

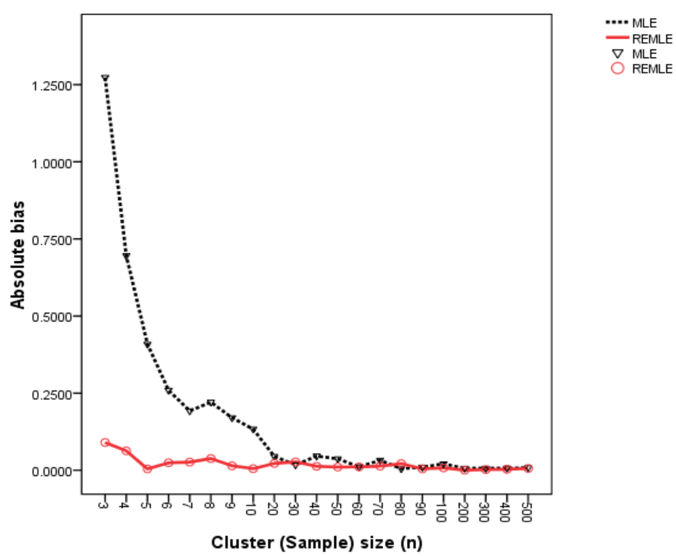


Figure 15: Plots of the absolute bias of ML and REML estimates of σ_e^2 for Model 2 against all the cluster (sample) sizes n (up to n = 500).

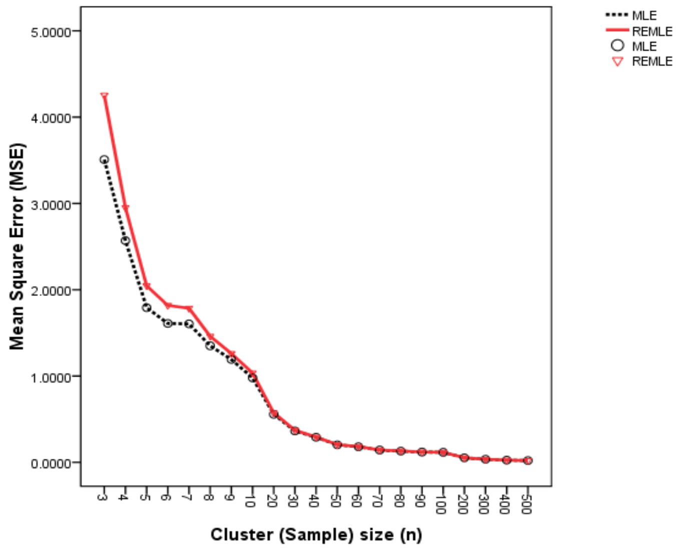


Figure 16: Plots of the MSE of ML and REML estimates of σ_e^2 for Model 2 against all the cluster (sample) sizes n (up to n = 500).

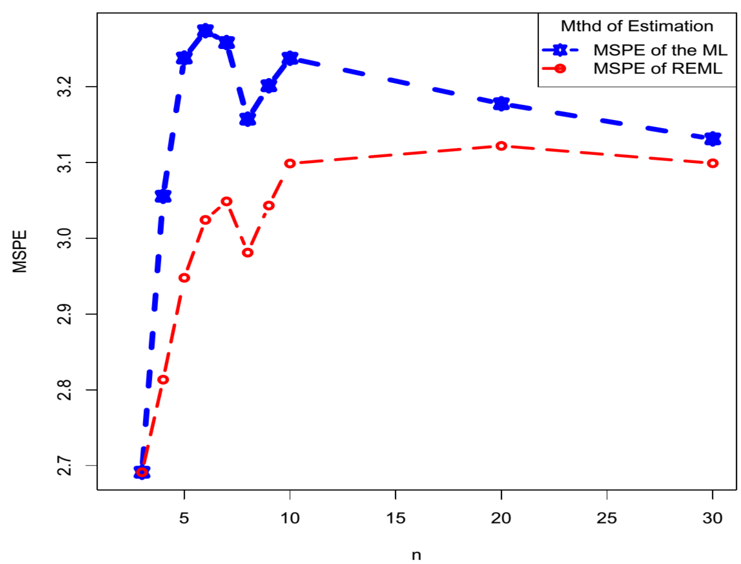


Figure 17: Plots of the MSPE against the cluster sizes n up to n = 30 based on ML and REML estimators for Model 2.

3.3. **Monte-Carlo Results for Model 3.** Model 3 has one random effect and eight fixed effects. Here, four other fixed effects parameters ($\beta_4 = 5.1, \beta_5 = -5.8, \beta_6 = 2.1, \beta_7 = -3.5$) were included in Model 2 to form Model 3 while

leaving the remaining terms and parameters as they were in Model 2. The fixed effects parameters showed the same behaviour as described in the case of the first and second models. One unique feature in Model 3 is that the minimum number of cluster allowable is 10 due to the number of parameters in the model.
Simulation results for Model 3 which contains eight fixed effects parameters are summarized in Table 6 and 7. The graphs of the estimated values of σ_b^2 by ML and REML methods are provided in Figure 18 for cluster sizes n up to n = 50. The plot of these estimates by both methods over all the cluster sizes employed is presented as Figure C.1 in Appendix C. The graphs of their respective bias and MSE of estimates are provided by Figure 19 and 20.

Table 6: Results of ML and REML estimations of the variance component σ_b^2 of the random effects parameter in Model 3. The average estimates of σ_b^2 using ML and REML estimators, their respective estimated bias and MSE over 1000 iterations are reported in the table for different number of cluster (n).

Number of clusters (n)	σ_b^2 (5.3)		Absolute Bias		MSE	
	ML	REML	ML	REML	ML	REML
10	1.6091	5.2868	3.6909	0.0132	16.6719	20.0315
15	2.4494	5.4195	2.8507	0.1195	10.9983	10.0848
20	3.1050	5.2876	2.1950	0.0124	7.4138	6.1011
30	3.9268	5.4140	1.3732	0.1140	4.1134	3.7823
40	4.2602	5.3661	1.0398	0.0661	2.8988	2.6727
50	4.3893	5.2592	0.9108	0.0408	2.2171	1.8743
60	4.5831	5.3143	0.7169	0.0143	1.7333	1.5592
70	4.6620	5.2864	0.6380	0.0136	1.3669	1.1827
80	4.7072	5.2500	0.5928	0.0500	1.2511	1.0810
90	4.7975	5.2822	0.5025	0.0178	1.1128	1.0103
100	4.8991	5.3394	0.4010	0.0394	0.9631	0.9283
200	5.0788	5.2977	0.2212	0.0023	0.4036	0.3807
300	5.1372	5.2827	0.1628	0.0173	0.3112	0.2986
400	5.2014	5.3110	0.0986	0.0110	0.2130	0.2107
500	5.1901	5.2773	0.1099	0.0227	0.1670	0.1598

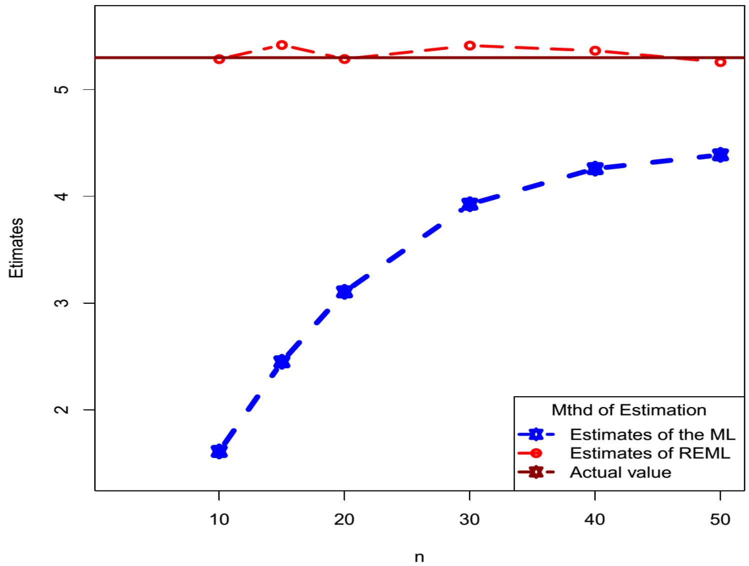


Figure 18: Plots of the estimated σ_b^2 for Model 3 using ML and REML methods against the cluster sizes n up to n = 50.

Table 7: Results of ML and REML estimations of the variance component σ_b^2 of the random effects parameter in Model 3. The average estimates of σ_b^2 using ML and REML estimators, their respective estimated bias and MSE over 1000 iterations are reported in the table for different number of cluster (n).

Number of clusters (n)	σ_e^2 (4)		Absolute Bias		MSE	
	ML	REML	ML	REML	ML	REML
10	3.8046	3.9987	0.1954	0.0013	0.9926	1.0775
15	3.9394	4.0365	0.0606	0.0365	0.6844	0.7223
20	3.9179	3.9844	0.0821	0.0156	0.5303	0.5418
30	3.9854	4.0302	0.0146	0.0302	0.3484	0.3570
40	3.9705	4.0039	0.0295	0.0039	0.2602	0.2638
50	3.9838	4.0105	0.0162	0.0105	0.2120	0.2147
60	3.9811	4.0034	0.0189	0.0034	0.1776	0.1793
70	4.0016	4.0207	0.0016	0.0207	0.1548	0.1568
80	3.9889	4.0055	0.0112	0.0055	0.1359	0.1369
90	3.9725	3.9872	0.0275	0.0128	0.1222	0.1225
100	3.9831	3.9965	0.0169	0.0036	0.1046	0.1050
200	4.0053	4.0120	0.0053	0.0120	0.0520	0.0523
300	4.0041	4.0085	0.0041	0.0085	0.0382	0.0383
400	3.9997	4.0030	0.0003	0.0030	0.0273	0.0273
500	3.9929	3.9956	0.0071	0.0044	0.0202	0.0202

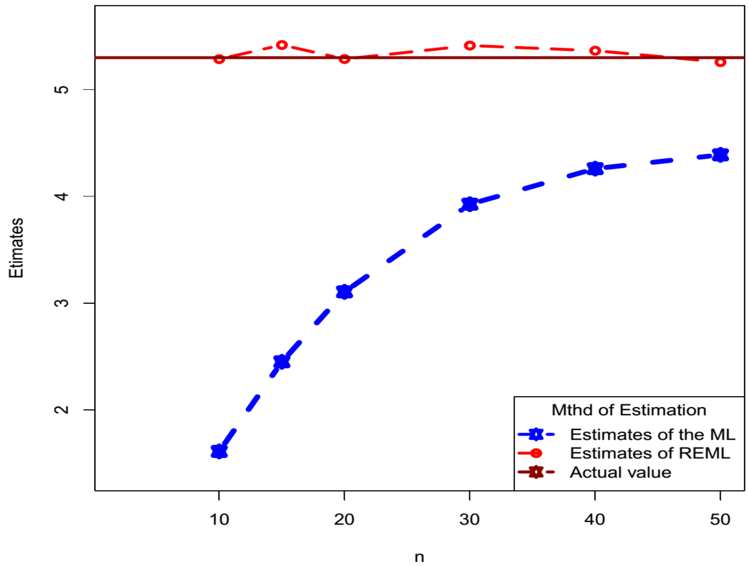


Figure 19: Plots of the absolute bias of ML and REML estimates of σ_b^2 for Model 3 against all the cluster (sample) sizes n (up to n = 500).

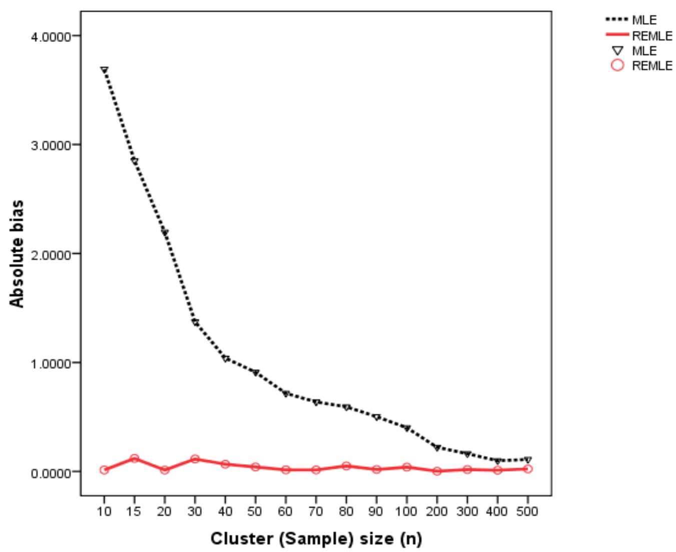
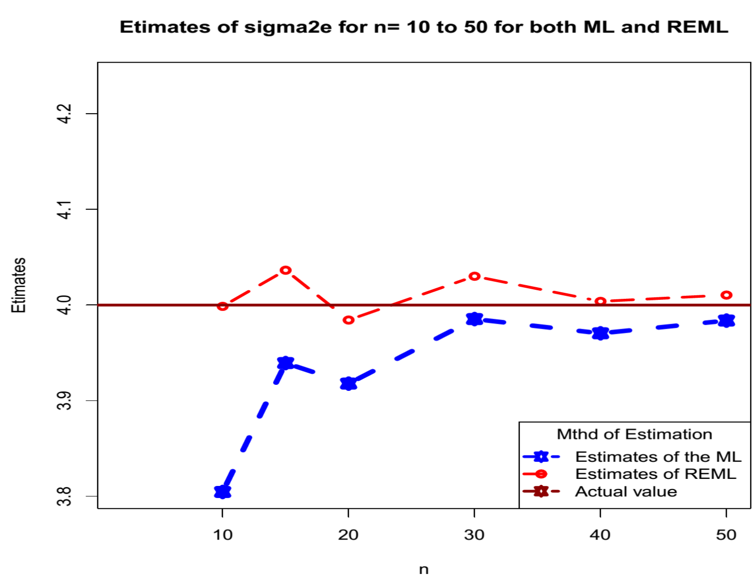


Figure 20: Plots of the MSE of ML and REML estimates of σ_b^2 for Model 3 against all the cluster (sample) sizes n (up to n = 500).

Similarly, the graphs of the estimated σ_e^2 by both ML and REML estimators are provided by Figure 21 for cluster sizes n up to n = 50. The plot of these estimates by both methods over all the cluster sizes employed is presented as Figure C.2 in Appendix C. The graphs of the absolute biases and MSEs of the two estimators for σ_e^2 are equally provided as Figures 22 and 23 respectively. The results show that the ML and REML estimates of $\sigma_b^2$and σ_b^2 behaved in a very similar manner like the first two models earlier discussed.



[htp]

Figure 21: Plots of the estimated σ_e^2 for Model 3 using ML and REML methods against the cluster sizes n up to n = 50.

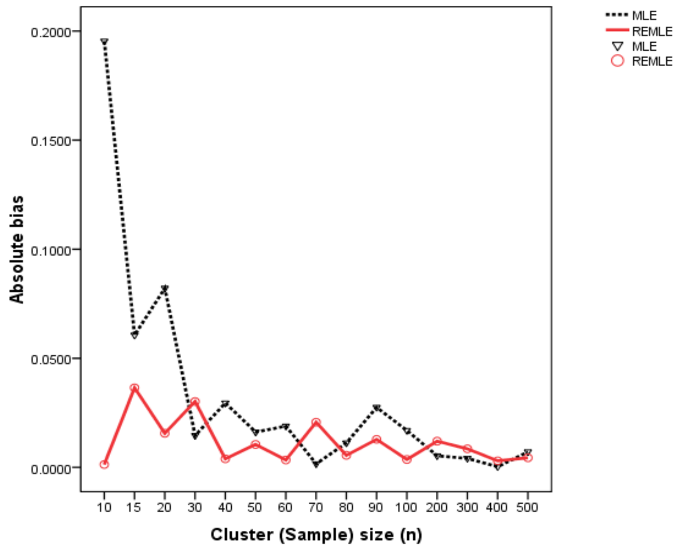


Figure 22: Plots of the absolute bias of ML and REML estimates of σ_e^2 for Model 3 against all the cluster (sample) sizes n (up to n = 500).

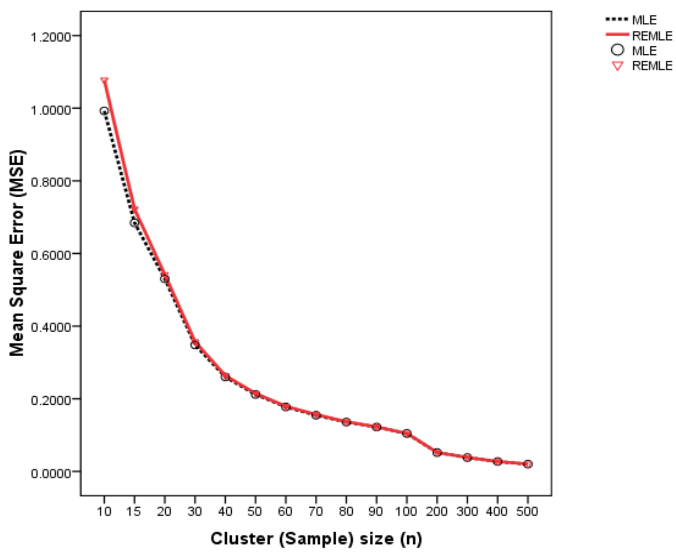


Figure 23: Plots of the MSE of ML and REML estimates of σ_e^2 for Model 3 against all the cluster (sample) sizes n (up to n = 500).

But, unlike the cases of Model 1 and Model 2, the MSEs of the REML estimation of σ_b^2 are lower than those of the ML method when the sample size is 15 or more. This can be observed in Figures 18, 19, 20 and 21. In this case, the biasness of ML for σ_b^2 has increased considerably compared to the first and second model. The estimated biases of REML for σ_b^2 seems not to follow a particular trend across the cluster sizes.
The behaviour of the estimates of σ_b^2 is once again very similar to those of the first and second model as shown by Figure 21. The plot of the MSPEs over the cluster sizes n up to n = 50 as shown in Figure 24 equally revealed that models fitted using the REML method have lower MSPEs than those fitted by the ML method. The plot of the MSPE over all the entire cluster sizes for Model 3 is presented as Figure C.3 in Appendix C.

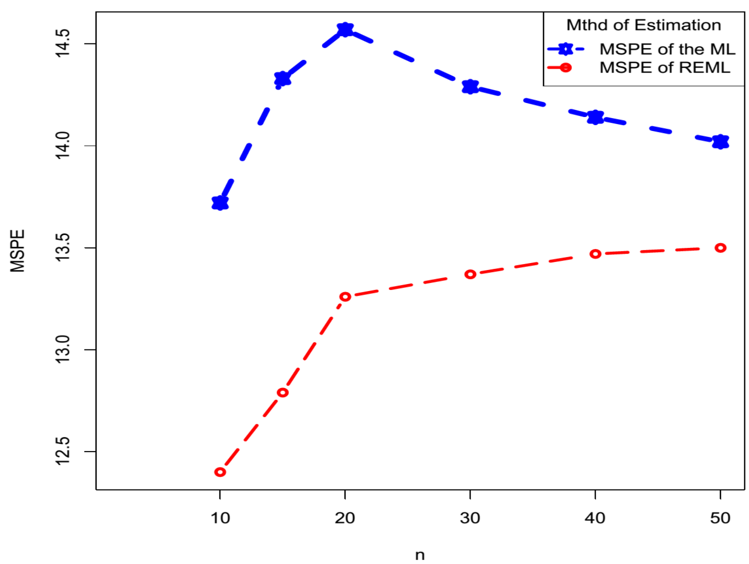


Figure 24: Plots of the MSPE against the cluster sizes n up to n = 50 based on ML and REML estimators for Model 3.

Table 8: The absolute biases of σ_b^2 and σ_e^2 in the three models at $n = 10$.

Parameter	Estimation Method	Model		
		Model 1	**Model 2**	**Model 3**
$\sigma_b^2(5.3)$	ML	1.2386	1.8307	3.6909
	REML	0.0199	0.0305	0.0132
$\sigma_e^2(4)$	ML	0.1752	0.1333	0.1954
	REML	0.0409	0.0052	0.0013

Table 8 presents the biases of the ML and REML methods for the three models at sample size (number of clusters) $n = 10$. The results showed that the bias of ML estimate of σ_b^2 increases considerably as the number of fixed effects parameters in the model increases. This is not the case for the REML method. On the other hand, the estimates of σ_e^2 do not seem to follow a particular trend but the bias in model 3 is slightly more than that of model 1. Therefore, results here show that the biasness of the ML estimation of σ_e^2 may not be seriously affected by the number of fixed effects parameters in the model.

4. DISCUSSION OF RESULTS

The focus of this work is to characterize the performances of ML and REML methods of estimating the variance components of linear mixed effects models. Three models with different number of fixed effects parameters were considered. The Monte-Carlo experiment was set such that the first, second and third models contained three, four and eight fixed effects parameters respectively. The simulated models mimicked typical models for longitudinal studies in real life cases where the number of observations per experimental unit or observational unit could be 4.

Data were simulated for the three models at different numbers observational unit (clusters) ranging from 3 to 500. The ML and REML methods were used to fit the models repeating the basic simulation one thousand times. Results of the study show that both methods (ML and REML) provided the same estimates of the fixed effects parameters. This implies that both methods are equally efficient at estimating the fixed effect parameters. It was further discovered, as can be seen in the various plots that the two methods are both biased for small samples, but as the sample size increases the estimates approached the true value asymptotically.

Furthermore, the results revealed that the ML estimates of the variance component of random effects σ_b^2 which is a vital component of the LME models are relatively more biased than the REML estimates especially for small sample size situations. The ML estimates of the variance component of the error terms are also biased. Results from the various plots clearly showed that, though the bias of the ML estimates are always higher than that of the REML estimates, the bias reduces as the sample size increases. Hence, when the sample size is very large, both methods provided almost the same estimates for the variance component of random effects and that of the error terms with bias close to zero (see Figures 8, 12 and 15).

Astonishingly, the results in Tables 2 to 5 showed that the mean square errors of the REML estimates of σ_b^2 and σ_e^2 are slightly higher than those of ML method especially in small sample size cases (see Figures 6, 9, 13 and 16). The only exception was observed in the case of Model 3 where the MSEs of the REML method were lower than those of the ML method (see Figures 20 and 22). This is an indication that the MSE of the REML method compared to that of the ML method may reduce as the number of fixed effects parameters increase. However, the MSE for both methods almost converge at large sample sizes.

Using the MSPE criterion, the results showed that models fitted by REML method yielded MSPEs that are relatively lower than those provided by the ML estimator. This indicates that models fitted by REML method have better fits than models fitted using the ML method. Nonetheless, as the sample size increases, the MSPEs of models from both methods were quite close (see Figures A.7, B.3 and C.3 in the Appendix).

Another important result from this study was that, the bias of the ML estimates increases as the number of fixed effects parameters in the model increases. This implies that the bias of the ML estimator is proportional to the number of fixed effects parameters in the model. When the sample size is 10 for instance, the bias of σ_b^2 for Model 1 with 3 fixed effects parameters is 1.2386. This bias increased to 1.8307 for Model 2 with 4 fixed effects parameters and increased to 4.2421 for Model 3 with 8 fixed effects parameters. The ML estimator is biased because it does not account for the loss of degrees of freedom for estimating the fixed effects parameters in the model. This makes it biased when there are many fixed effects parameters in the model, substantiating the findings of Haville (1977). However, results here show that the biasness of the ML estimation of the variance component of the error terms σ_e^2 may not necessarily depend on the number of fixed effects parameters in the model.

5. Conclusion

Two methods of estimating a classical linear mixed effects model ML and REML methods - have been presented and examined in this work. Extensive Monte-Carlo studies were carried out to establish the relative efficiencies of these methods under varying parameter combinations.

Based on the findings from this work, it was observed that the choice of which method to adopt depends on a number of factors such as number of fixed effects parameters in the model and the sample size. It can however, be recommended that the REML should be preferred to ML if there are quite a number of fixed effects parameters in the model. However, when the sample size is fairly large (for example n = 200), the two methods demonstrated apparent similar efficiency based on the adopted assessment criteria in this work.

It is worth being mentioned that the ML and REML methods of parameter estimation were only considered for estimating mixed effects model in this work. The strength of other methods of estimation such as the Minimum Norm Quadratic Unbiased Estimator (MINQUE) as reported by Rao (1971), Hendersons estimation procedures (Henderson, 1953) and Bayesian techniques (Gelfand et. al, 1990) need to be investigated empirically for their relative efficiencies under

varying conditions.
Mixed effects models are very powerful, flexible, and besides, very useful in many real life situations like in Medicine, Engineering and humanities. Therefore, further advancement and developments concerning this concept should constantly be undertaken for better improvement. Applications of this kind of model using some of its identified modeling techniques shall be presented in future work.

References

[1] orbeil, R.R. and Searle, S.R. (1976). A comparison of variance component estimators. Biometrics 32, 779-791.
[2] empster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Ser. B 39, 1-38.
[3] elfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990): Illustration of Bayesian inference in normal data models using Gibbs sampling, Journal of the American Statistical Association, 85(412): 972985.
[4] artley and Rao (1967): Maximum Likelihood Estimation for the mixed analysis of variance model. Biometrika, 54(1): 93-108.
[5] arville, D. A. (1974). Bayesian inference for variance components using only error contrasts, Biometrika, 61: 383385.
[6] arville, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems, Journal of the American Statistical Association 72: 320338.
[7] enderson CR. (1953): Estimation of Variance and Covariance Components. Biometrics; 9(2): 223-52.
[8] aird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. Biometrics, 38: 963974
[9] indstrom, M. J. and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. Journal of the American Statistical Association 83: 10141022.
[10] ongford, N. T. (1993). Random Coefficient Models, Oxford University Press, New York.
[11] cLachlan, G.J. and Krishnan, T. (1997). The EM-Algorithm and Extension. New York: Wiley.
[12] inheiro, J.C. and Bates, D.M. (2000). Mixed-Effects Models in S and S-PLUS. Springer, New York.
[13] ao, C.R. (1971): Minimum Variance Quadratic Unbiased Estimation of Variance Components. Journal of Multivariate Analysis, 1(4): 445-56.
[14] earle, S. R., Casella, G. and McCulloch, C. E. (1992). Variance Components, Wiley, New York.
[15] histed, R. A. (1988). Elements of Statistical Computing, Chapman & Hall, London.
[16] erbeke, G. and Molenberghs, G. (2001). Linear Mixed Models for Longitudinal Data. New York: Springer.
[17] ang, W. and Heckman, N. (2009). Identifiability in linear mixed models. Technical report, Department of Statistics, University of British Columbia.
[18] olfinger, R., Tobias, R. and Sall, J. (1991). Mixed models: A future direction, SUGI 16: 13801388.

[19] u, C.F.J. (1983). On the convergence properties of the EM algorithm. Annals of Statistics 11, 95-103.

[20] u, L. (2010). Mixed effects models for complex data, Monographs on statistics and applied probability 113, University of British Columbia Vancouver, Canada.

6. APPENDIXES

6.1. **APPENDIX A.** THE GRAPHS OF THE RESULTS OF ML AND REML ESTIMATORS OVER LARGE SAMPLES (CLUSTER SIZES) FOR MODEL 1.

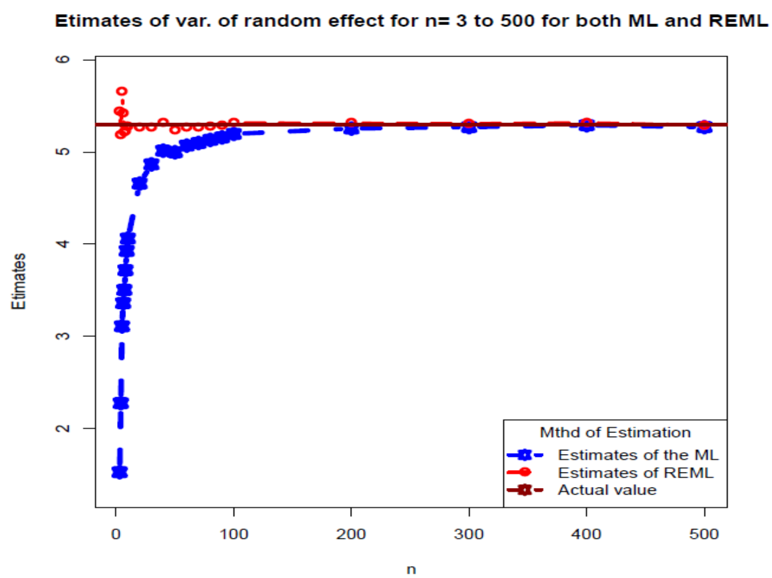


Figure A.1: Plots of the estimated σ_b^2 for Model 1 using the ML and REML methods over all the cluster sizes n up to n = 500.

6.2. **APPENDIX B.** THE GRAPHS OF THE RESULTS OF ML AND REML ESTIMATORS OVER LARGE SAMPLES (CLUSTER SIZES) FOR MODEL 2.

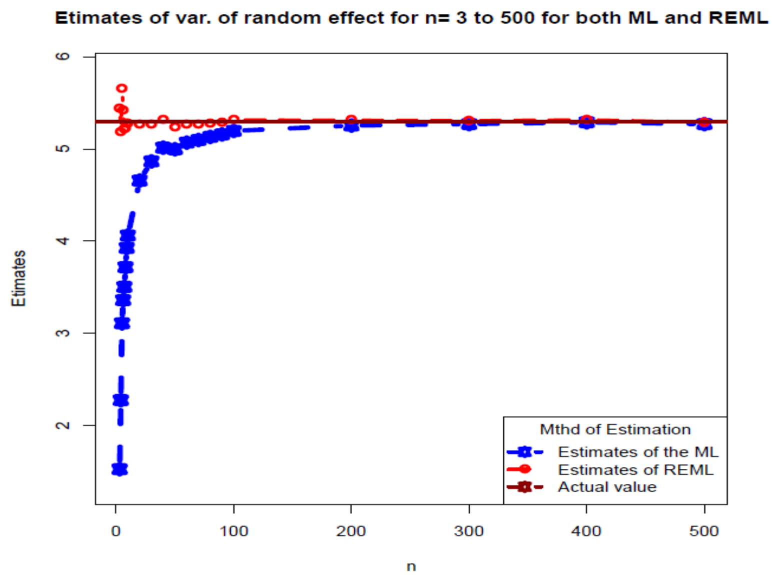


Figure A.1: Plots of the estimated σ_b^2 for Model 1 using the ML and REML methods over all the cluster sizes n up to n = 500.

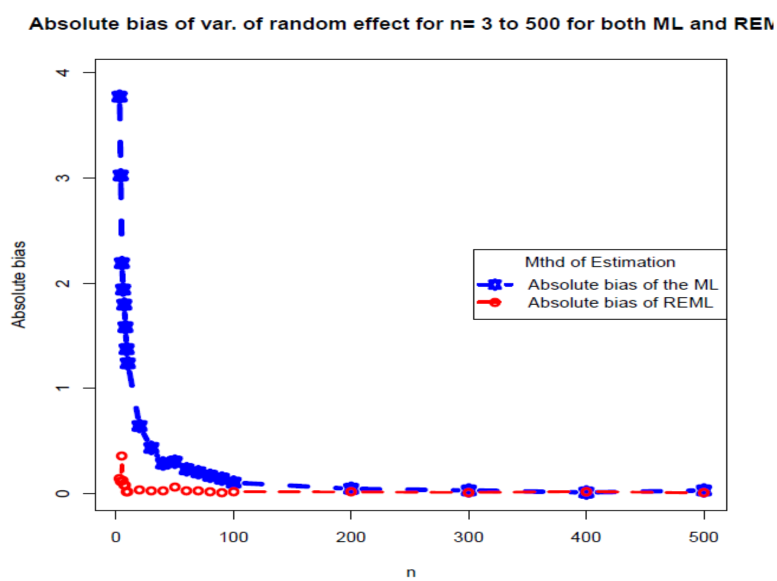


Figure A.2: Plots of the absolute bias of ML and REML estimates of σ_b^2 for Model 1 against all the cluster sizes n up to n = 500.

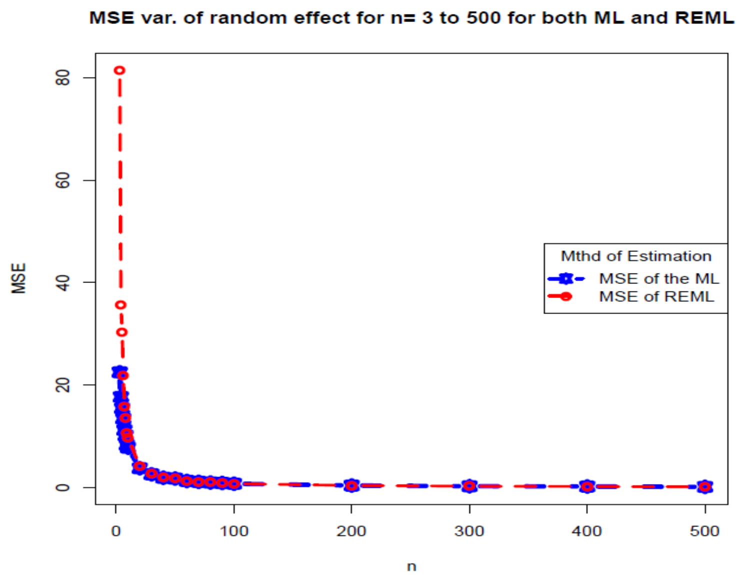


Figure A.3: Plots of the MSE of ML and REML estimates of σ_b^2 for Model 1 against all the cluster sizes n up to n = 500.

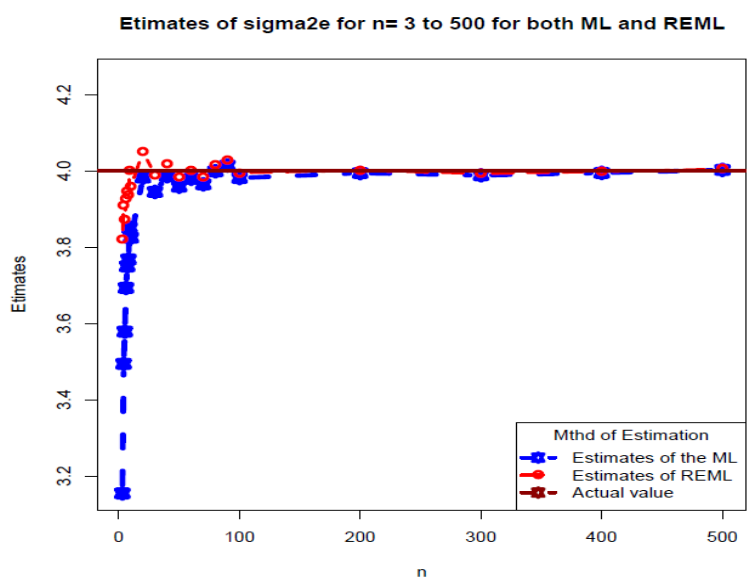


Figure A.4: Plots of the estimated σ_e^2 for Model 1 using the ML and REML methods over all the cluster sizes n up to n = 500.

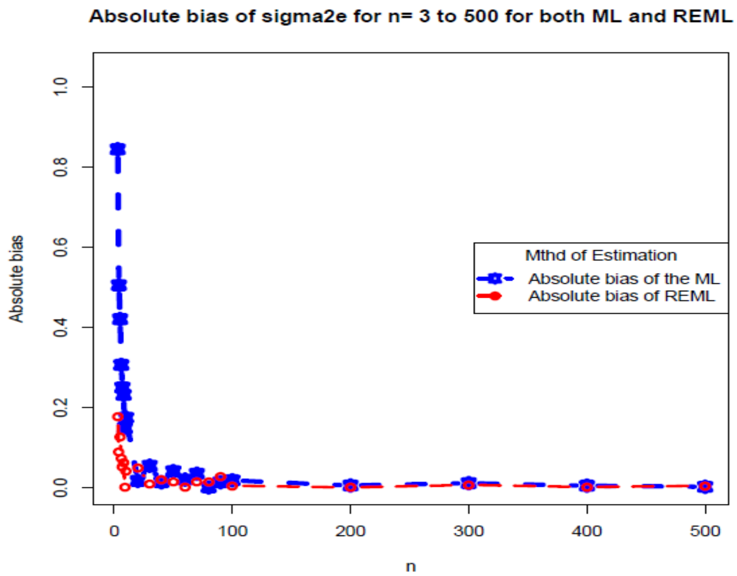


Figure A.5: Plots of the absolute bias of ML and REML estimates of σ_e^2 for Model 1 against all the cluster sizes n up to n = 500.

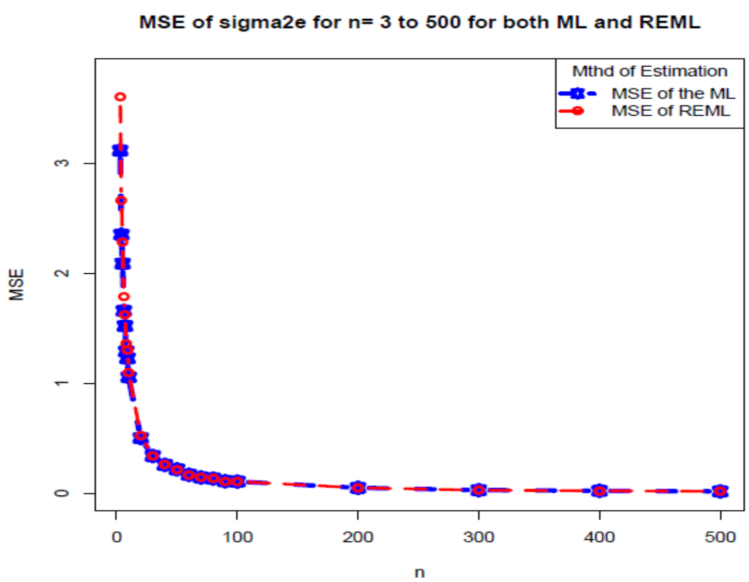


Figure A.6: Plots of the MSE of ML and REML estimates of σ_e^2 for Model 1 against all the cluster sizes n up to n = 500.

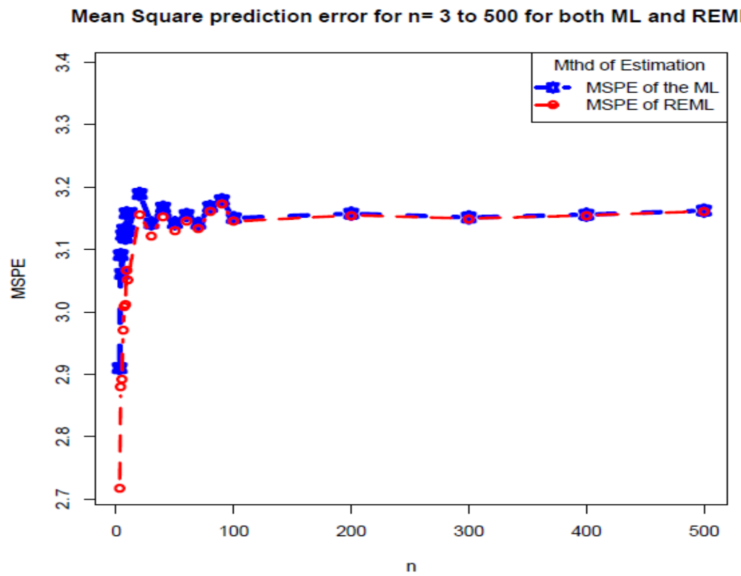


Figure A.7: Plots of the MSPE against all the cluster sizes n up to n = 500 based on ML and REML estimators for Model 1.

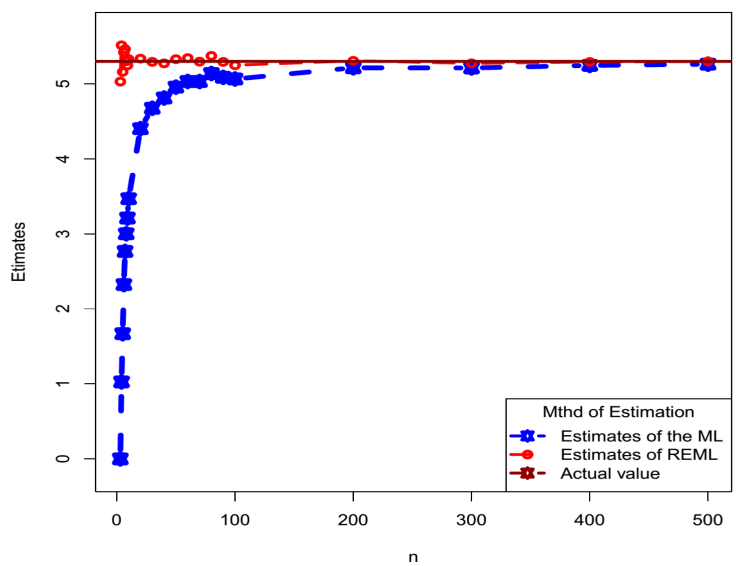


Figure B.1: Plots of the estimated σ_b^2 for Model 2 using the ML and REML methods over all the cluster sizes n up to n = 500.

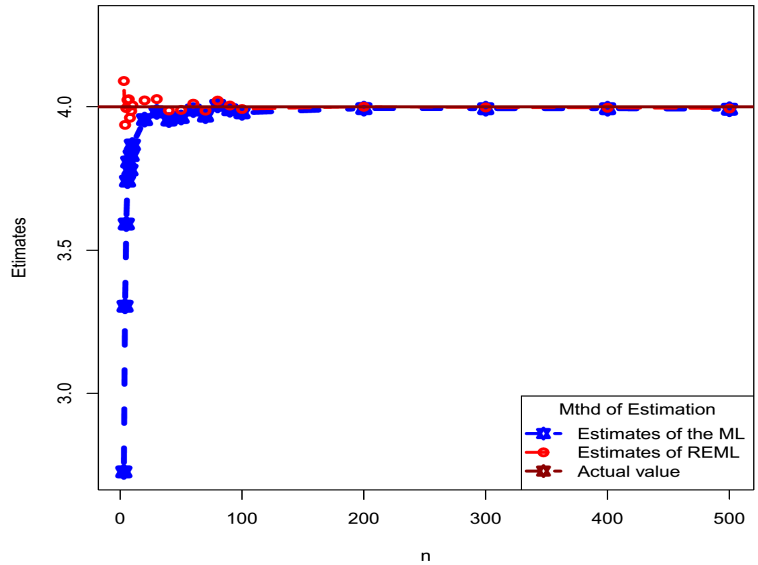


Figure B.2: Plots of the estimated σ_e^2 for Model 2 using the ML and REML methods over all the cluster sizes n up to n = 500.

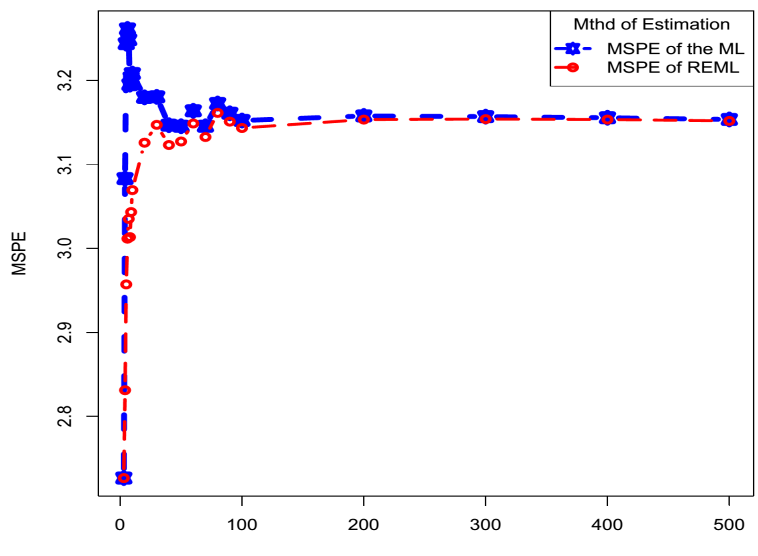


Figure B.3: Plots of the MSPE against all the cluster sizes n up to n = 500 based on ML and REML estimators for Model 2.

6.3. **APPENDIX C.** THE GRAPHS OF THE RESULTS OF ML AND REML ESTIMATORS OVER LARGE SAMPLES (CLUSTER SIZES) FOR MODEL 3.

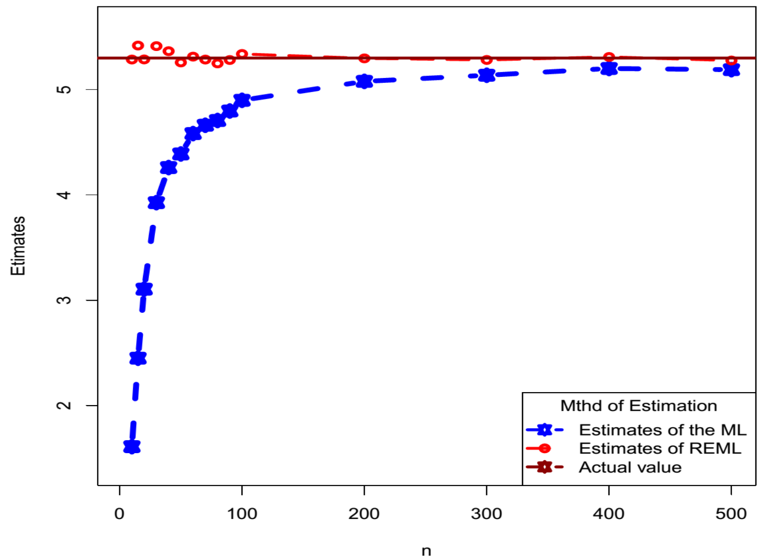


Figure C.1: Plots of the estimated σ_b^2 for Model 3 using the ML and REML methods over all the cluster sizes n up to n = 500.

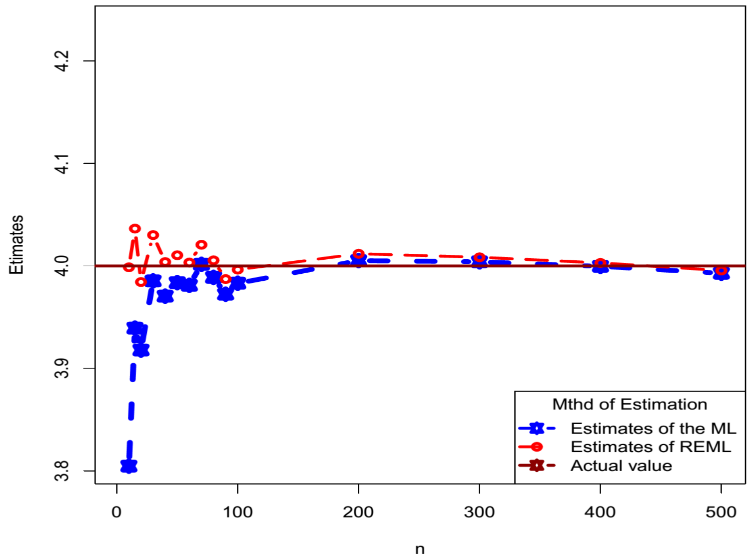


Figure C.2: Plots of the estimated σ_e^2 for Model 3 using the ML and REML methods over all the cluster sizes n up to n = 500.

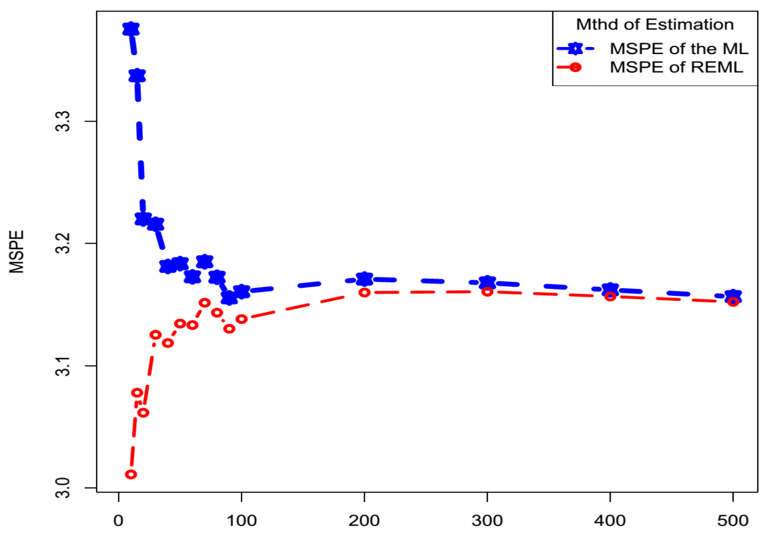


Figure C.3: Plots of the MSPE against all the cluster sizes n up to n = 500 based on ML and REML estimators for Model 3.