Evaluation of a New Variance Components Estimation Method—Modified Henderson's Method 3 With the Application of Two Way Mixed Model

Author: Weigang Qie; Chenfan Xu
Supervisor: Lars Rönnegård
June 10th, 2009

D-level Essay in Statistics, Spring 2009
Department of Economics and Society, Dalarna University College.

Evaluation of a New Variance Components Estimation Method—Modified Henderson's Method 3 With the Application of Two Way Mixed Model

Weigang Qie; Chenfan Xu

June 10, 2009

Abstract

A two-way linear mixed model with three variance components as σ_1^2, σ_2^2 and σ_e^2 is applied to evaluate the performance of modified Henderson's method 3 developed by Al-Sarraj and Rosen (2007). The focus of modified procedure is on the estimation of σ_1^2 which variance components is mainly concerned. The modified estimator is expected to perform better than unmodified Henderson's method 3 in terms of MSE. But it also follows the demerits of unmodified one, i.e. lost uniqueness, negative estimates. The criteria used to show the performance of modified estimator compared with unmodified one, ML and REML are bias, MSE and the probability of getting negative estimate. Al-Sarraj and Rosen (2007) suggested us to divide the estimation of σ_1^2 of Henderson's method 3 and its modified into Partition I and Partition II. One way to solve the problem of lost unique estimators is to compare the MSE of Partition I and II, then select the one with smaller MSE. The performances of these estimators in terms of MSE are shown by the means of simulations. MSE effects of imbalance and number of observations are given. Based on the MSE comparison of Partition I and II, there should exist a boundary value of σ_2^2 to favor Partition I, otherwise II. From the effects of σ_2^2 and σ_1^2 to MSE, a 'small' values range of $\sigma_2^2 < 0.1$ is recommended to prefer to the Partition I of Henderson's method 3 and its modified compared with Partition II. Then, a ratio range of $\sigma_2^2/\sigma_1^2 < 1.0$ is obtained for wide application. Modified Henderson's method 3 has achieved substantially improvement over unmodified one in terms of MSE, as well as the probability of getting negative estimate. It is also computationally faster than ML and REML and may for some cases performs better in terms of MSE. The split-plot design experiment application shows us that the modified estimator can improve unmodified one.

Keywords: Variance components, Modified Henderson's method 3, MSE, Monte Carlo simulation.

Contents

Notation list

MSE	Mean Square Errors
SSR	Reduction in sum of squares
SST	Total sum of squares
SSE	Residual error sum of squares
REML	Restricted Maximum Likelihood
ML	Maximum Likelihood
n	Obersvations
N	Sample size (number of simulations)
p	Levels in u_1
q	Levels in u_2
b	Numbers of fixed effects
σ^2	Variance components
$\hat{\sigma}^2_{u1}$	Estimator of Partition I for Henderson's method 3
$\hat{\sigma}^2_{1}$	Estimator of Partition II for Henderson's method 3
$\hat{\sigma}^2_{11}$	Estimator of Partition I for modified Henderson's method 3
$\hat{\sigma}^2_{12}$	Estimator of Partition II for modified Henderson's method 3
$\hat{\sigma}^2_{u1REML}$	Estimator of REML
$\hat{\sigma}^2_{u1ML}$	Estimator of ML

1 Introduction

1.1 Background

Variance components estimation has a wide application, i.e. genetics, pharmacy and econometrics. The model applied is a kind of hierarchical linear model assuming a hierarchy of different populations which yields random effects. It is reasonable to add random effects to classical linear model which includes fixed effects only. McCulloch and Searle (2002) provided a decision tree to assist us to decide whether the parameters are fixed or not. The rule is that if we can reasonably assume the levels of the factor come from a probability distribution, then treat the factor as random; otherwise fixed. The likelihood ratio test to decide whether the random effects exist or not was introduced in Giampami and Singer (2009). If the model contains both fixed and random effects, we can extend classical model to mixed linear model which is commonly used.

Inquiring for an appropriate method to estimate variance components has attached much attention in statistical research in different experiments. The most commonly used method for balanced data is analysis of variance (ANOVA) which equates the observed mean squares to their expected values and the variance components estimates are obtained by the solving these equations. Graybill and Hultquist (1961) illustrated that ANOVA estimators were the best quadratic unbiased estimators (BQUE) and has minimum variance among other unbiased estimators with the quadratic functions of observations. The ANOVA estimator could get negative estimates which may cause terrible problems to analyze. In general case, the data are often unbalanced. As long as the ANOVA being used in unbalanced data, their good properties except unbiasedness of this estimator are lost. Rao (1972) introduced a method called Minimum Variance Quadratic Unbiased Estimation (MIVQUE). A *priori* values must be supplied before the application of MIVQUE. Only if perfect priori values equaling to the true values of the variance components are given, this estimator will achieve minimum sample variance. For a one-way classification random model under normality with σ^2_a and σ^2_e, MIVQUE used to estimate σ^2_a often has much smaller variance than the usual ANOVA estimator and they differ a little based on numerical results; see Swallow and Searle (1978). The applications of Maximum Likelihood (ML) together with its comparison with Restricted Maximum Likelihood REML based on some algorithms were described in Harville (1977). ML approaches are used to estimate variance components by maximizing the likelihood over the positive space of the variance components parameters. Some of attractive features and deficiencies for ML are given, i.e. takes no account of the loss in degrees of freedom resulting form estimating the fixed effects. Restricted maximum likelihood (REML) was developed by Patterson and Thompson (1971) to modified ML which considers the loss of freedom degrees and corrects the bias of ML. Many of iterative algorithms such as Newton-Raphson and Fisher score are used for the REML and ML variance components estimation. We can not expect that a single numerical computing process yields a prefect estimate both form REML and ML. The converge rate, computational requirements and special properties of experiments are seen as important rules to find appropriate algorithms. As a limitation of the ML and REML estimators, the experiments with large observations may cause computational problem calculated by iterative algorithms.

Three well known Henderson's methods to solve difficulty with unbalanced data for estimating variance components are developed by Henderson (1953). All the three are adaptations of the ANOVA method of equating analysis of variance sums of squares to their expected values. The estimators are unbiased, but they also have demerits, i.e. negative estimates, different solutions yielded from the different set of equations for the same parameter; see Searle, Casella and McCulloch (1992). Al-Sarraj and Rosen (2007) modified the Henderson's method 3 by relax the unbiasedness to improve it in terms of MSE. The estimator obtained from the new method is expected to have smaller MSE than unmodified one. That is where we shall test via the means of simulations in the article. The performane of the new modified estimator compared with ML and REML should also be considered.

There are no perfect estimators in all experiments with the applications of these methods referred above. Several estimators applied to practical data set can produce substantially different results. Christensen, Pearson and Johnson (1992) showed examples that the values of estimates yielded by the ANOVA, ML and REML are uncommonly different. So some criteria are in need to evaluate the performance of the different estimators. Generally, the unbiased estimators are required because its good properties, i.e. closest to the true value when sample size is large. Corbeil and Searle (1976) considered the mean squared errors (MSE) as one of the criterion. The MSE[1] which includes both the dispersion and deviation degrees for an estimator is a measure to quantify the distance between estimates and true values. It is a function of sample variance and bias for the estimators. The unbiased estimator with smallest MSE performs better than other estimators. But, sometimes the biased estimators may have a smaller MSE than the unbiased ones. According to the definition of sample distribution for the estimators, the rules to prefer which kind of estimators are derived. Since the unbiased estimators are closer to the true values in this situation; if the experiments are repeated for many times the unbiased estimators with larger MSE are favored over the biased estimators with smaller MSE. Otherwise, if the experiments took place only once or repeated few times, the biased estimators with smaller MSE are preferred. Moreover, Kelly and Mathew (1994) recommended that the explicit analytic expressions with easy computation for estimators is considered. Since the estimates of variance components should be positive according to its definition, the probability of getting negative estimate is also seen as a measure to show the difference among the estimators. The noniterative estimators with explicit expression unlike ML and REML, i.e. mainly concerned estimators of $\hat{\sigma}^2_{u1}$ and $\hat{\sigma}^2_{11}$, are compared together with $\hat{\sigma}^2_{1}$, $\hat{\sigma}^2_{12}$, $\hat{\sigma}^2_{u1REML}$ and $\hat{\sigma}^2_{u1ML}$ in terms of these criteria described above.

1.2 Aim and Outline of the Article

The aim of the article is to evaluate modified Henderson's method 3 with the application of two-way linear mixed model by the means of simulations compared with unmodified one, REML and ML. As a new method obtained from the Henderson's method 3, the modified estimator is expected to achieve some improvements over the unmodified one. Moreover, this new method is a noniterative estimator which should be favored over iterative estimators i.e. ML and REML. It is necessary and meaningful to show its performance by comparison with the other estimators, especially the unmodified one. The criteria to evaluate are given in subsection 1.1. The MSE is considered as the main concern because of its wide application and good properties, i.e. often used with aim of comparison between different estimators, and includes both the effects of variance and bias.

In section 1, a simple introduction about the variance components estimations is first given. This section also states the aim and proposes the mixed model used in our article. The methods of unmodified and modified Henderson's method 3 together with ML and REML are described in section 2. The process and results of Monte Carlo comparison are shown in section 3. In section 3, the differences between examples are described by the measure of imbalance. We also recommend which situation is the modified Henderson's method 3 favored over the other estimators. Furthermore, in section 4, the Henderson's method 3 and its modified, ML and REML are implemented to apply the Split-Plot design experiment. The results also show the modified estimator perform well compared with unmodified one. Based on the analysis simulation and data application results, the conclusion in section 5 is drawn that modified Henderson's method 3 can be suggested as the appropriate estimator in terms of MSE. Finally the limitations of the modified Henderson's method 3 are described in section 6.

[1]The definitions of the bias and MSE are given in APPENDIX B

2 Methodology

2.1 Two-Way Mixed Linear Mixed Model

We consider the two-way mixed model in matrix form:

$$Y = X\beta + Z_1 u_1 + Z_2 u_2 + e \tag{1}$$

where $Y_{n\times 1}$ is the observation vector and distributed as a multivariate normal $MVN(X\beta, V)$ with $V = \sigma_1^2 Z_1' Z_1 + \sigma_2^2 Z_2' Z_2 + \sigma_e^2$, $V_1 = Z_1' Z_1$ and $V_2 = Z_2' Z_2$ are also defined. $X_{n\times 1}$ is the full column rank design matrix for fixed effects, $Z_{1(n\times p)}$ and $Z_{1(n\times p)}$ are design matrices for random effects, **e** is the error term which is distributed as multivariate $e \sim MVN(0, \sigma_e^2 I)$. β is the fixed effects, u_1 and u_2 with p and q levels are the random effects which are distributed as multivariate $u_1 \sim MVN(0, \sigma_1^2 I)$, $u_2 \sim MVN(0, \sigma_2^2 I)$ respectively. Let us define $\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_e^2)'$ which is so called variance components. The σ_1^2 is only interested because the modified procedure is focus on the estimation of this variance components. Then six different estimators of σ_1^2 are proposed in our article. We calculate the biases, probability of getting negative estimate and MSE of to evaluate modified Henderson's method 3 by the comparison with the others.

2.2 Henderson's Method 3

The method named Henderson's Method 3 is first established by Henderson (1953). Together with it, another two methods, the Henderson's Method 1 and Henderson's Method 2 are also derived. The differences of them lie in the quadratic forms and experiments application. If the three of Henderson's methods apply to the balanced data, their estimates are the same as each other. The Henderson's Method 3 is focused on the issue of variance component estimation for unbalanced data. The core procedures are to solve the equations of the reductions in sums of squares of the quadratic forms and their expectations. Its advantages include no strong distribution assumption, and unbiased estimator as well. And the demerits can be noticed in the aspects of negative estimates and no unique estimators which is caused by the no unique set of decompositions of the reductions in sums of squares to estimate. In order to solve the problem of lost unique estimators, Al-Sarraj and Rosen (2007) suggested us to divide decompositions used to estimate into Partition I with three variance components and Partition II with two variance components respectively. So Partition I and II are compared in terms of MSE. Then the one has smaller MSE would be selected as the appropriate estimator, otherwise the other. The Partition I or II with smaller MSE can also be chosen to modify.

2.2.1 Variance Components Estimator for Partition I

The theory of reductions in sums of squares is introduced by Searle (1987). Let $R(\cdot)$ denotes the reductions in sums of squares which is equal to the SSR of some linear models. For the one-way random model $y_{ij} = \mu + \alpha_i + e_{ij}$ where i is the level of random effects α and j is the observations of each i, the difference of $R(\mu, \alpha) - R(\mu)$ interprets the reductions in sums of squares due to fitting to the random effect α after μ that is already considered. Hence, let us define the notation $R(\cdot/\cdot)$ to denote the difference of the reductions in sums of squares between the different models. The $R(\cdot)$ and $R(\cdot/\cdot)$ are distributed as non-central χ^2 under the normality assumption. Searle (1987) also showed these reductions in sums of squares and their differences are independent of each other and of SSE.

The submodels of full model (1) used to obtain estimation equations in Al-Sarraj and Rosen (2007) are given as:

$$\begin{gathered} Y = X\beta + e \ \text{ for } R(\beta) \\ Y = X\beta + Z_1 u_1 + e \ \text{ for } R(\beta, u_1) \\ Y = X\beta + Z_2 u_2 + e \ \text{ for } R(\beta, u_2) \end{gathered}$$

There are two sets of estimation equations can be considered because of three elements.

$$\left\{ \begin{matrix} R(u_1/\beta) \\ R(u_2/\beta, u_1) \\ SSE \end{matrix} \right\} \quad \text{or} \quad \left\{ \begin{matrix} R(u_1/\beta) \\ R(u_1/u_2, \beta) \\ SSE \end{matrix} \right\}$$

where the SSE denotes the residual error sum of squares.

Define the projection matrix as $P_\omega = \omega\left(\omega'\omega\right)^{-}\omega'$ which is idempotent[2] matrix. Hence, the first set of the above equations is suggested by Al-Sarraj and Rosen (2007) to estimate the Partition I of Henderson's method 3 and the following of projection matrices for estimation are proposed.

$$P_x = X\left(X'X\right)^{-}X$$
$$P_{x1} = (X, Z_1)\left((X, Z_1)'(X, Z_1)\right)^{-}(X, Z_1)'$$
$$P_{x12} = (X, Z_1, Z_2)\left((X, Z_1, Z_2)'(X, Z_1, Z_2)\right)^{-}(X, Z_1, Z_2)'$$

By using the projection matrices given above, the differences of reductions in sums of squares $R(\cdot/\cdot)$ used to equate their expectations are:

$$R(u_1/\beta) = R(\beta, u_1) - R(\beta) = Y'(P_{x1} - P_x)Y$$
$$R(u_2/\beta, u_1) = R(\beta, u_1, u_2) - R(\beta, u_1) = Y'(P_{x12} - P_{x1})Y$$
$$SSE = Y'Y - R(\beta, u_1, u_2) = Y'(I - P_{x12})Y$$

Their expectations are presented below:

$$E\begin{bmatrix} Y'(P_{x1} - P_x)Y \\ Y'(P_{x12} - P_{x1})Y \\ Y'(I - P_{x12})Y \end{bmatrix} = J\begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_e^2 \end{bmatrix} \tag{2}$$

where $J = \begin{bmatrix} tr(P_{x1} - P_x)V_1 & tr(P_{x1} - P_x)V_2 & tr(P_{x1} - P_x) \\ tr(P_{x12} - P_{x1})V_1 & tr(P_{x12} - P_{x1})V_2 & tr(P_{x12} - P_{x1}) \\ tr(I - P_{x12})V_1 & tr(I - P_{x12})V_2 & tr(I - P_{x12}) \end{bmatrix}$

Since $P_{x1}V_1 = V_1$, $P_{x12}V_2 = V_2$ and $P_{x12}V_1 = V_1$ where V_1 and V_2 are defined in subsection 2.1, the simple form of J is

$$J = \begin{bmatrix} tr((P_{x1} - P_x)V_1) & tr((P_{x1} - P_x)V_2) & tr(P_{x1} - P_x) \\ 0 & tr((P_{x12} - P_{x1})V_2) & tr(P_{x12} - P_{x1}) \\ 0 & 0 & tr(I - P_{x12}) \end{bmatrix}$$

Here let us define some notations to simplify to express

$$\begin{aligned} A &= (P_{x1} - P_x),\ B = (P_{x12} - P_{x1}), C = (I - P_{x12}), \\ a &= tr((P_{x1} - P_x)V_1), b = tr((P_{x12} - P_{x1})V_2),\ c = tr(I - P_{x12}) \end{aligned}$$

$$d = tr((P_{x1} - P_x)V_2), e = tr((P_{x12} - P_{x1})V_1), f = tr(P_{x1} - P_x) \tag{3}$$

Here $\hat{\sigma}_{u1}^2$ is denoted the estimator of σ_1^2 for Partition I of Henderson's method 3. Then by solving the equations in (2), the estimates of variance components are

$$\begin{bmatrix} \hat{\sigma}_{u1}^2 \\ \hat{\sigma}_2^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = J^{-1}\begin{bmatrix} Y'(P_{x1} - P_x)Y \\ Y'(P_{x12} - P_{x1})Y \\ Y'(I - P_{x12})Y \end{bmatrix} \tag{4}$$

Thus the expression of $\hat{\sigma}_{u1}^2$ with simple form is:

$$\hat{\sigma}_{u1}^2 = \frac{Y'AY}{a} - \frac{d\left(Y'BY\right)}{ab} + \frac{k\left(Y'CY\right)}{abc} \tag{5}$$

[2]Matrix A satisfies $AA = A$, it can be seen as a idempotent matrix

where $k = d \times e - f \times b$ and the notations are defined in (3).
Hence, the sample variance of $\hat{\sigma}^2_{u1}$ is calculated as:

$$\begin{aligned} D\left(\hat{\sigma}^2_{u1}\right) &= \left[\tfrac{2}{a^3} tr\left(AV_1AV_1\right)\right]\sigma^4_1 \\ &+ \left[\tfrac{2}{a^2} tr\left(AV_2AV_2\right) + \tfrac{2d^2}{a^2b^2} tr\left(BV_2BV_2\right)\right]\sigma^4_2 \\ &+ \left[\tfrac{4}{a^2} tr\left(AV_1AV_2\right)\right]\sigma^2_1\sigma^2_2 + \left[\tfrac{4}{a^2} tr\left(AV_1A\right)\right]\sigma^2_1\sigma^2_e \\ &+ \left[\tfrac{4}{a^2} tr\left(AV_2A\right) + \tfrac{4d^2}{a^2b^2} tr\left(BV_2B\right)\right]\sigma^2_2\sigma^2_e \\ &+ \left[\tfrac{2}{a^2} tr\left(AA\right) + \tfrac{2d^2}{a^2b^2} tr\left(BB\right) + \tfrac{2k^2}{a^2b^2c^2} tr\left(CC\right)\right]\sigma^4_e \end{aligned} \tag{6}$$

where the notations are the same as in (3).

Since $\hat{\sigma}^2_{u1}$ is an unbiased estimator, so the predicted MSE of $\hat{\sigma}^2_{u1}$ is $MSE\left(\hat{\sigma}^2_{u1}\right) = D\left(\hat{\sigma}^2_{u1}\right)$. From the equation (6), $MSE\left(\hat{\sigma}^2_{u1}\right)$ includes six terms and depends on σ^2_1, σ^2_2 and σ^2_e.

2.2.2 Variance Components Estimator for Partition II

There are more sets of equations for estimation than variance components. In order to solve this problem, Al-Sarraj and Rosen (2007) developed the variance components estimator for Partition II to estimate σ^2_1 with different set based on the model (1). The MSE of partition II is also calculated. We compare the MSE of Partition I and II, and then select the one with smaller MSE to modify.

Then the projection matrix used to estimate the Partition II is:

$$P_{x2} = (X, Z_2)\left((X, Z_2)'(X, Z_2)\right)^{-}(X, Z_2)'$$

The set of estimation equations for the Partition II is given:

$$\left\{ \begin{array}{c} R\left(u_1/\beta, u_2\right) \\ SSE \end{array} \right\}$$

Where $R\left(u_1/\beta, u_2\right) = R\left(\beta, u_1, u_2\right) - R\left(\beta, u_2\right)$
$= Y'\left(P_{x12} - P_{x2}\right)Y$
and $SSE = Y'Y - R\left(\beta, u_1, u_2\right) = Y'\left(I - P_{x12}\right)Y$
The expectation of equations used to estimate partition II of are

$$E\left[\begin{array}{c} Y'\left(P_{x12} - P_{x2}\right)Y \\ Y'\left(I - P_{x12}\right)Y \end{array}\right] = K\left[\begin{array}{c} \sigma^2_1 \\ \sigma^2_e \end{array}\right] \tag{7}$$

where $K = \left[\begin{array}{cc} tr\left(\left(P_{x12} - P_{x2}\right)V_1\right) & tr\left(P_{x12} - P_{x2}\right) \\ 0 & \left(I - P_{x12}\right) \end{array}\right]$
Some notations are defined to simplify:

$$E = P_{x12} - P_{x2}, g = tr\left(\left(P_{x12} - P_{x2}\right)V_1\right), l = tr\left(P_{x12} - P_{x2}\right) \tag{8}$$

Here $\hat{\sigma}^{2}_1$[3] is denoted the estimator for Partition I of Henderson's method 3. Then by solving the equations in (7), the estimates of variance components are

$$\left[\begin{array}{c} \hat{\sigma}^2_1 \\ \hat{\sigma}^2_e \end{array}\right] = K^{-1}\left[\begin{array}{c} Y'\left(P_{x12} - P_{x2}\right)Y \\ Y'\left(I - P_{x12}\right)Y \end{array}\right] \tag{9}$$

Thus, the expression of estimator for Partition II $\hat{\sigma}^2_1$ is:

[3] The estimator $\hat{\sigma}^2_1$ can be obtained from the reduced model method which is discussed in APPENDIX D

$$\hat{\sigma}_1^2 = \frac{1}{g} Y' EY - \frac{l}{cg} Y' CY \tag{10}$$

where the notations are used in (3) and (8)
The sample variance of $\hat{\sigma}_1^2$ is calculated:

$$\begin{aligned} D\left(\hat{\sigma}_1^2\right) &= \left[\tfrac{2tr(EV_1EV_1)}{g^2}\right] \sigma_1^4 \\ &+ \left[\tfrac{4tr(EV_1E)}{g^2}\right] \sigma_1^2\sigma_e^2 \\ &+ \left[\tfrac{2tr(EE)}{g^2} + \tfrac{2l^2}{g^2c}\right] \sigma_e^4 \end{aligned} \tag{11}$$

Because of its unbiasedness, so the MSE of $\hat{\sigma}_1^2$ is $MSE\left(\hat{\sigma}_1^2\right) = D\left(\hat{\sigma}_1^2\right)$. From the equation (14), $MSE\left(\hat{\sigma}_1^2\right)$ includes three terms and depends on σ_1^2 and σ_e^2. Variance components σ_2^2 does not effect $MSE\left(\hat{\sigma}_1^2\right)$.

$MSE\left(\hat{\sigma}_{u1}^2\right)$ and $MSE\left(\hat{\sigma}_1^2\right)$ Comparison It is obvious to see the difference of (6) and (11). The equation (6) includes the terms of σ_2^4, $\sigma_1^2\sigma_2^2$ and $\sigma_2^2\sigma_e^2$ which (14) does not have. If the σ_1^2 and σ_e^2 are fixed, there should exist a boundary value of σ_2^2 which make $MSE\left(\hat{\sigma}_{u1}^2\right) = MSE\left(\hat{\sigma}_1^2\right)$. There is a ascending trend of $MSE\left(\hat{\sigma}_{u1}^2\right)$ for increasing σ_2^2. Hence, if $\hat{\sigma}_{u1}^2$ is conerned, a 'small' values range of σ_2^2 which makes $MSE\left(\hat{\sigma}_{u1}^2\right) < MSE\left(\hat{\sigma}_1^2\right)$ can be obtained to prefer to $\hat{\sigma}_{u1}^2$ in terms of MSE. The 'small' values range of σ_2^2 to favor $\hat{\sigma}_{u1}^2$ is confirmed by the means of simulations in section 3.

2.3 Modified Henderson's Method 3

Here we summarize the theory of modified Henderson's method 3 developed by Al-Sarraj and Rosen (2007). It is applied to improve the estimation equations of Henderson's method 3 by multiplying some constants. These constants to modify Henderson' method 3 are determined by minimizing the coefficients of leading terms in its MSE, i.e. σ_1^4 and σ_e^2. The modified estimator relaxes unbiasedness caused by the constants, but it should perform better than unmodified one in terms of MSE. It also has no unique estimators and is divided in to Partition I and II which are similar with the unmodified estimators $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_1^2$.

2.3.1 Modified Variance Components Estimator for Partition I

Here $\hat{\sigma}_{11}^2$ denotes Partition I of modified Henderson's method 3. $\hat{\sigma}_{11}^2$ is modified from the Partition I of unmodified estimator $\hat{\sigma}_{u1}^2$. Based on the set of equations (4), a new class of equations is presented:

$$E\begin{bmatrix} c_1 Y' \left(P_{x1} - P_x\right) Y \\ c_1 d_1 Y' \left(P_{x12} - P_{x1}\right) Y \\ c_1 d_2 Y' \left(I - P_{x12}\right) Y \end{bmatrix} = J \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_e^2 \end{bmatrix} \tag{12}$$

Where J is the same as in equation (2), and $c_1 \geq 0$, d_1 and d_2 are defined as the constants to be determined by minimizing the leading terms of MSE of $\hat{\sigma}_{11}^2$.

By solving equation (12), we have the expression of variance components estimation.

$$\begin{bmatrix} \hat{\sigma}_{11}^2 \\ \hat{\sigma}_2^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = J^{-1} \begin{bmatrix} c_1 Y' \left(P_{x1} - P_x\right) Y \\ c_1 d_1 Y' \left(P_{x12} - P_{x1}\right) Y \\ c_1 d_2 Y' \left(I - P_{x12}\right) Y \end{bmatrix} \tag{13}$$

The expression of $\hat{\sigma}_{11}^2$ is obtained from equation (13):

$$\hat{\sigma}_{11}^2 = \frac{c_1 Y' AY}{a} - \frac{c_1 d_1 d \left(Y' BY\right)}{ab} + \frac{c_1 d_2 k \left(Y' CY\right)}{abc} \tag{14}$$

Where the notations are the same as in (3)
The sample variance of $\hat{\sigma}^2_{11}$ is

$$\begin{aligned} D\left(\hat{\sigma}^2_{11}\right) = &\left[\tfrac{2c_1^2}{a^3} tr\left(AV_1AV_1\right)\right]\sigma_1^4 \\ &+\left[\tfrac{2c_1^2}{a^2} tr\left(AV_2AV_2\right)+\tfrac{2c_1^2d_1^2d^2}{a^2b^2} tr\left(BV_2BV_2\right)\right]\sigma_2^4 \\ &+\left[\tfrac{4c_1^2}{a^2} tr\left(AV_1AV_2\right)\right]\sigma_1^2\sigma_2^2+\left[\tfrac{4c_1^2}{a^2} tr\left(AV_1A\right)\right]\sigma_1^2\sigma_e^2 \\ &+\left[\tfrac{4c_1^2}{a^2} tr\left(AV_2A\right)+\tfrac{4c_1^2d_1^2d^2}{a^2b^2} tr\left(BV_2B\right)\right]\sigma_2^2\sigma_e^2 \\ &+\left[\tfrac{2c_1^2}{a^2} tr\left(AA\right)+\tfrac{2c_1^2d_1^2d^2}{a^2b^2} tr\left(BB\right)+\tfrac{2c_1^2d_2^2k^2}{a^2b^2c^2} tr\left(CC\right)\right]\sigma_e^4 \end{aligned} \tag{15}$$

Since unbiasedness is lost, we calculate the expectation of $\hat{\sigma}^2_{11}$:

$$\begin{aligned} E\left(\hat{\sigma}^2_{11}\right) = &\left(\tfrac{c_1}{a} tr\left(AV_1\right)\right)\sigma_1^2 \\ &+\left(\tfrac{c_2}{a} tr\left(AV_2\right)-\tfrac{c_1d_1d}{ab} tr\left(BV_2\right)\right)\sigma_2^2 \\ &+\left(\tfrac{c_1}{a} tr\left(A\right)-\tfrac{c_1d_1d}{ab} tr\left(B\right)+\tfrac{c_1kd_2}{abc} tr\left(C\right)\right)\sigma_e^2 \end{aligned} \tag{16}$$

The bias of $\hat{\sigma}^2_{11}$ is obtained from equation (15).

$$\begin{aligned} Bias\left(\hat{\sigma}^2_{11}\right) &= E\left(\hat{\sigma}^2_{11}\right)-\sigma_1^2 \\ &=\left(\tfrac{c_1}{a} tr\left(AV_1\right)-1\right)\sigma_1^2 \\ &+\left(\tfrac{c_2}{a} tr\left(AV_2\right)-\tfrac{c_1d_1d}{ab} tr\left(BV_2\right)\right)\sigma_2^2 \\ &+\left(\tfrac{c_1}{a} tr\left(A\right)-\tfrac{c_1d_1d}{ab} tr\left(B\right)+\tfrac{c_1kd_2}{abc} tr\left(C\right)\right)\sigma_e^2 \end{aligned} \tag{17}$$

Thus, based on equations of (14) and (16), the MSE of $\hat{\sigma}^2_{11}$ is:

$$\begin{aligned} MSE\left(\hat{\sigma}^2_{11}\right) &= D\left(\hat{\sigma}^2_{11}\right)+Bias^2\left(\hat{\sigma}^2_{11}\right) \\ &=\left[\tfrac{2c_1^2}{a^3} tr\left(AV_1AV_1\right)+\left(c_1-1\right)^2\right]\sigma_1^4 \\ &+\left[\tfrac{2c_1^2}{a^2} tr\left(AV_2AV_2\right)+\tfrac{2c_1^2d_1^2d^2}{a^2b^2} tr\left(BV_2BV_2\right)+r^2\right]\sigma_2^4 \\ &+\left[\tfrac{4c_1^2}{a^2} tr\left(AV_1AV_2\right)+2\left(c_1-1\right)r\right]\sigma_1^2\sigma_2^2 \\ &+\left[\tfrac{4c_1^2}{a^2} tr\left(AV_1A\right)+2\left(c_1-1\right)t\right]\sigma_1^2\sigma_e^2 \\ &+\left[\tfrac{4c_1^2}{a^2} tr\left(AV_2A\right)+\tfrac{4c_1^2d_1^2d^2}{a^2b^2} tr\left(BV_2B\right)+2rt\right]\sigma_2^2\sigma_e^2 \\ &+\left[\tfrac{2c_1^2}{a^2} tr\left(AA\right)+\tfrac{2c_1^2d_1^2d^2}{a^2b^2} tr\left(BB\right)+\tfrac{2c_1^2d_2^2k^2}{a^2b^2c^2} tr\left(CC\right)+t^2\right]\sigma_e^4 \end{aligned} \tag{18}$$

with $r=\frac{c_1d}{a}-\frac{dc_1d_1}{a}$ and $t=\frac{c_1}{a} tr\left(A\right)-\frac{dc_1d_1}{ab} tr\left(B\right)+\frac{c_1kd_2}{ab}$

In order to achieve expectation results that $MSE\left(\hat{\sigma}^2_{11}\right)\leq MSE\left(\hat{\sigma}^2_{u1}\right)$, we need to obtain appropriate values of constants used in equation (13). Based on several steps of comparison with the coefficients of σ_1^4, σ_2^4 and σ_e^4 of $MSE\left(\hat{\sigma}^2_{11}\right)$, Al-Sarraj and Rosen (2007) gave us the results of constants:

$$c_1=\frac{1}{\frac{2}{a^2} tr\left(AV_1AV_1\right)+1} \tag{19}$$

$$d_1 = \frac{1}{\frac{2}{b^2} tr\left(BV_2BV_2\right) + 1} \tag{20}$$

$$d_2 = \frac{\frac{d}{b} d_1 tr\left(B\right) - tr\left(A\right)}{\left(\frac{k}{b}\right)\left(\frac{2}{c}+1\right)} \tag{21}$$

The above three constants have been verified that they minimize the coefficients terms of σ_1^4, σ_2^4 and σ_e^2 respectively in equation (18). Then the coefficients of the three terms are smaller than the same terms respectively in equation (6). Moreover, there are three remaining cross terms corresponding to $\sigma_1^2\sigma_2^2$, $\sigma_1^2\sigma_e^2$ and $\sigma_2^2\sigma_e^2$ in (18) need to compare with the same terms in (6).

Two conditions corresponding to cross terms of $MSE\left(\hat{\sigma}_{11}^2\right)$ must be satisfied to have the remaining cross terms smaller are established by Al-Sarraj and Rosen (2007).

Condition 1 $tr\left(A\right) \leq \frac{d}{b} d_1 tr\left(B\right)$ *and* $tr\left(A\right) \geq \frac{d}{b} tr\left(B\right) - \frac{(2+c)(1+c_1)}{c_1}$

Condition 2 $tr\left(A\right) > \frac{d}{b} d_1 tr\left(B\right)$ *and* $d_1 = 1$

After the constants in (19), (20) and (21) are estimated, if one of the conditions given above is satisfied, we have the $MSE\left(\hat{\sigma}_{11}^2\right) \leq MSE\left(\hat{\sigma}_{u1}^2\right)$. Then $\hat{\sigma}_{u1}^2$ can be reasonable to modify to $\hat{\sigma}_{11}^2$ in terms of MSE.

2.3.2 Modified Variance Components Estimator for Partition II

Here $\hat{\sigma}_{12}^2$[4] is defined as the Partition II of modified Henderson's method 3. The set of equations to solve $\hat{\sigma}_{11}^2$ is similar with $\hat{\sigma}_1^2$.

$$E\begin{bmatrix} c_2 Y'\left(P_{x12} - P_{x2}\right) Y \\ c_2\epsilon_1 Y'\left(I - P_{x12}\right) Y \end{bmatrix} = K\begin{bmatrix} \sigma_1^2 \\ \sigma_e^2 \end{bmatrix} \tag{22}$$

where the constants c_2, ϵ_1 to modify $\hat{\sigma}_1^2$ are determined by minimizing the leading terms of $MSE\left(\hat{\sigma}_{12}^2\right)$, i.e. σ_1^4 and σ_e^4.

The expression of the variance components estimations is given by solving equation (22).

$$\begin{bmatrix} \hat{\sigma}_{12}^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = K^{-1}\begin{bmatrix} c_2 Y'\left(P_{x12} - P_{x2}\right) Y \\ c_2\epsilon_1 Y'\left(I - P_{x12}\right) Y \end{bmatrix} \tag{23}$$

So, the estimator $\hat{\sigma}_{12}^2$ is obtained from equation (23):

$$\hat{\sigma}_{12}^2 = \frac{c_2}{g} Y'EY - \frac{c_2\epsilon_1 l}{cg} Y'CY \tag{24}$$

The sample variance of $\hat{\sigma}_{12}^2$ is:

$$\begin{aligned} D\left(\hat{\sigma}_{12}^2\right) &= \left[\frac{2c_2^2 tr(EV_1EV_1)}{g^2}\right]\sigma_1^4 \\ &+ \left[\frac{4c_2^2 tr(EV_1E)}{g^2}\right]\sigma_1^2\sigma_e^2 \\ &+ \left[\frac{2c_2^2 tr(EE)}{g^2} + \frac{2c_2^2\epsilon_1^2 l^2}{g^2 c}\right]\sigma_e^4 \end{aligned} \tag{25}$$

Then the bias of $\hat{\sigma}_{12}^2$ is calculated as:

[4] The estimator similar with $\hat{\sigma}_1^2$ can also be obtained from the reduced model method.

$$Bias\left(\hat{\sigma}_{12}^2\right) = (c_2 - 1)\,\sigma_1^2 + \left(\frac{c_2 l}{g} - \frac{c_2 \epsilon_1 l}{g}\right)\sigma_e^2 \tag{26}$$

Based on equations (25) and (26) the $MSE\left(\hat{\sigma}_{12}^2\right)$ is given:

$$\begin{aligned} MSE\left(\hat{\sigma}_{12}^2\right) &= D\left(\hat{\sigma}_{12}^2\right) + Bias^2\left(\hat{\sigma}_{12}^2\right) \\ &= \left[\tfrac{2c_2^2 tr(EV_1EV_1)}{g^2} + (c_2-1)^2\right]\sigma_1^4 \\ &+ \left[\tfrac{4c_2^2 tr(EV_1E)}{g^2} + 2\,(c_2-1)\,\tfrac{c_2 l}{g}\left(1-\tfrac{l}{g}\right)\right]\sigma_1^2\sigma_e^2 \\ &+ \left[\tfrac{2c_2^2 tr(EE)}{g^2} + \tfrac{2c_2^2\epsilon_1^2 l^2}{g^2 c} + \left(\tfrac{c_2 l}{g}\left(1-\tfrac{l}{g}\right)\right)^2\right]\sigma_e^4 \end{aligned} \tag{27}$$

In order to achieve the expectation result that $MSE\left(\hat{\sigma}_{12}^2\right) \leq MSE\left(\hat{\sigma}_1^2\right)$. The contants of c_2 and $\epsilon_1$are also obtained by minimizing the coefficients of σ_1^4 and σ_e^4 involving the leading terms in $MSE\left(\hat{\sigma}_{12}^2\right)$. The results suggested from Kelly and Mathew (1994) are given in (28) and (29) respectively.

$$c_2 = \frac{1}{\frac{2}{g^2}\left(EV_1EV_1\right) + 1} \tag{28}$$

$$\epsilon_1 = \frac{1}{\frac{2}{c} + 1} \tag{29}$$

It is verified that the two constants minimize the coefficients corresponding to σ_1^4 and σ_e^4 in (27). That means the coefficients of terms of and in (27) are smaller than the same terms in (11) respectively. Moreover, Al-Sarraj and Rosen (2007) suggested a condition which is satisfied to have the cross coefficients terms of $\sigma_1^2\sigma_e^2$ in (27) smaller than the same term in (9)

Condition 3 $tr\left(EV_1E\right) \geq \frac{2g(c_2-1)(c_2-\epsilon_1)}{4\left(1-c_2^2\right)}$

If the constants in (28) and (29) are estimated, and the above condition is satisfied, then $\hat{\sigma}_{12}^2$ is favored over $\hat{\sigma}_1^2$ in terms of MSE.

$MSE\left(\hat{\sigma}_{11}^2\right)$ and $MSE\left(\hat{\sigma}_{12}^2\right)$ Comparison The difference between $MSE\left(\hat{\sigma}_{11}^2\right)$ and $MSE\left(\hat{\sigma}_{12}^2\right)$ is similar with $MSE\left(\hat{\sigma}_{u1}^2\right)$ and $MSE\left(\hat{\sigma}_1^2\right)$. Hence, if $\hat{\sigma}_{11}^2$ is concerned, a 'small' values range of σ_2^2 can also be obtained by means of simulations to choose $\hat{\sigma}_{11}^2$ rather than $\hat{\sigma}_{12}^2$ in terms of MSE.

2.4 Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML)

2.4.1 Equations to Estimate $\hat{\sigma}_{u1ML}^2$ and $\hat{\sigma}_{u1REML}^2$

$\hat{\sigma}_{u1ML}^2$ is defined as the estimator of ML. For mixed model in (1), the log-likelihood function for ML is

$$\log L_{ML} = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|V| - \frac{1}{2}\left(Y - X\beta\right)' V^{-1}\left(Y - X\beta\right) \tag{30}$$

Then we take the first and second derivatives of the equation (30) with respect to β and variance components σ^2 respectively, Searle, Casella and McCulloch (1992) gave us the equations.

First:

$$\frac{\partial \log L_{ML}}{\partial \beta} = X'V^{-1}Y - X'V^{-1}X\beta \tag{31}$$

$$\frac{\partial \log L_{ML}}{\partial \sigma_i^2} = -\frac{1}{2} tr\left(V^{-1} Z_i Z_i^{'}\right) + \frac{1}{2}\left(Y - X\beta\right)^{'} V^{-1} Z_i Z_i^{'} V^{-1}\left(Y - X\beta\right) \tag{32}$$

Second:

$$\frac{\partial^2 \log L_{ML}}{\partial \sigma_i^2 \sigma_j^2} = -\frac{1}{2} tr\left(V^{-1} Z_i Z_i^{'} V^{-1} Z_j Z_j^{'}\right) - \frac{1}{2}\left(Y - X\beta\right)^{'} V^{-1} Z_i Z_i^{'} V^{-1} Z_j Z_j^{'}\left(Y - X\beta\right) \tag{33}$$

The elements of ML information martrix which is defined as $\mathcal{I}_{ML}$

$$-E\left(\frac{\partial^2 \log L_{ML}}{\partial \sigma_i^2 \sigma_j^2}\right) = \frac{1}{2} tr\left(V^{-1} Z_i Z_i^{'} V^{-1} Z_j Z_j^{'}\right) \tag{34}$$

with $i = j = 0, 1, 2, \sigma_0^2 = \sigma_e^2$ and $Z_0 Z_0^{'} = I$ (35)

Since there are nonlinear forms to estimate the elements of in (32) and (33), the solutions of ML are usually obtained by iterative algorithms.

$\hat{\sigma}^2_{u1REML}$[5] is defined as the estimator of REML. REML is an unbiased esatimtor modified from ML. For model (1), the log-likelihood function for the REML is

$$\log L_{REML} = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|V| - \frac{1}{2}\log\left|X^{'} V^{-1} X\right| - \frac{1}{2}\left(Y - X\beta\right)^{'} V^{-1}\left(Y - X\beta\right) \tag{36}$$

Similar with the ML approach, the derivatives to maximize (35) with respect to β and variance components σ^2 the equations are given by Harville (1977):

First:

$$\frac{\partial \log L_{REML}}{\partial \beta} = X^{'} V^{-1} Y - X^{'} V^{-1} X\beta \tag{37}$$

$$\frac{\partial \log L_{REML}}{\partial \sigma_i^2} = -\frac{1}{2} tr\left(P Z_i Z_i^{'}\right) \tag{38}$$

$$+\frac{1}{2}\left(Y - X\beta\right)^{'} V^{-1} Z_i Z_i^{'} V^{-1}\left(Y - X\beta\right)$$

Second:

$$\frac{\partial^2 \log L_{REML}}{\partial \sigma_i^2 \sigma_j^2} = -\frac{1}{2} tr\left(P\left(\partial^2 V/\partial^2 \sigma_i^2 \sigma_j^2 - Z_i Z_i^{'} P Z_j Z_j^{'}\right)\right) - \frac{1}{2}\left(Y - X\beta\right)^{'} V^{-1}\left(\partial^2 V/\partial^2 \sigma_i^2 \sigma_j^2 - 2 Z_i Z_i^{'} P Z_j Z_j^{'}\right) V^{-1}\left(Y - X\beta\right) \tag{39}$$

The elements of REML information matrix which is defined as $\mathcal{I}_{REML}$

$$-E\left(\frac{\partial^2 \log L_{REML}}{\partial \sigma_i^2 \sigma_j^2}\right) = \frac{1}{2} tr\left(P Z_i Z_i^{'} P Z_j Z_j^{'}\right) \tag{40}$$

where $P = V^{-1} - V^{-1} X\left(X^{'} V^{-1} X\right)^{'} X^{'} V^{-1}$ and the notations are the same as (35)

[5]We use lmer( ) function of lme4 package in R to estimate $\hat{\sigma}^2_{u1ML}$ and $\hat{\sigma}^2_{u1REML}$

2.4.2 Summary of Algorithms

The algorithms of the Newton-Raphson and the Fisher score are commonly used for ML and REML variance components estimation. We give a summary of the two algorithms. The application of iterative algorithms to estimate $\hat{\sigma}^2_{u1ML}$ and $\hat{\sigma}^2_{u1REML}$ are similar with each other.

Let $L\left(\sigma^2\right)$ be the likelihood of variance components σ^2 for ML or REML of model (1). The aim is to find the solution $\hat{\sigma}^2$ of when the $L\left(\sigma^2\right)$ is maxima.

A brief description of Newton-Raphson algorithm is given as follows.

The first gradient of $L\left(\sigma^2\right)$ with σ^2 is defined as $\triangledown\left(\sigma^2\right)$:

$$\triangledown\left(\sigma^2\right) = \left(\frac{\partial \log L\left(\sigma^2\right)}{\partial \sigma_1^2}, \frac{\partial \log L\left(\sigma^2\right)}{\partial \sigma_2^2}, \frac{\partial \log L\left(\sigma^2\right)}{\partial \sigma_e^2}\right) \tag{41}$$

Then the second of derivative of $L\left(\sigma^2\right)$ with σ^2 is denoted by H. Here the is an 3×3 symmetric matrix H with elements $h_{ij} = \frac{\partial^2 \log L\left(\sigma^2\right)}{\partial \sigma_i^2 \sigma_j^2}$ where i and j are defined in (35). Now the Taylor's second series of $\triangledown\left(\sigma^2\right)$ with the starting values $\sigma^2_{(0)}$ is:

$$\triangledown\left(\hat{\sigma}^2\right) = \triangledown\left(\sigma^2_{(0)}\right) + H_0\left(\sigma^2\right)\left(\sigma^2 - \hat{\sigma}^2\right) \tag{42}$$

If $\hat{\sigma}^2$ make the maximum of $L\left(\sigma^2\right)$, then $\triangledown\left(\hat{\sigma}^2\right) = 0$ which can be replaced in (42).

The solution of $\hat{\sigma}^2$ is:

$$\hat{\sigma}^2 = \sigma^2_{(0)} - H_0^{-1}\left(\sigma^2\right)\triangledown\left(\sigma^2_{(0)}\right) \tag{43}$$

After m^{th} iteration, the Newton-Raphson algorithm is:

$$\sigma^2_{(m+1)} = \sigma^2_{(m)} - H_m^{-1}\left(\sigma^2\right)\triangledown\left(\sigma^2_{(m)}\right) \tag{44}$$

Under the converge restriction which depends on the special requirements of real experiments, $\sigma^2_{(m+1)} \to \hat{\sigma}^2$ when $\triangledown\left(\sigma^2_{(m+1)}\right) \approx 0$.

Davidson (2003) introduce the Fisher's score algorithm which is similar with Newton-Raphson.

Let us define Fisher score $\mathcal{S}\left(\sigma^2\right)$ which is equal to $\triangledown\left(\sigma^2\right)$. By replacing the $-H_0\left(\sigma^2\right)$ with its expectation in (42) which is the so called information matrix denoted by $\mathcal{I}$. Hence we have the iterative solution of for Fisher score:

$$\mathcal{S}\left(\hat{\sigma}^2\right) = \mathcal{S}\left(\sigma^2_{(0)}\right) - \mathcal{I}_0\left(\sigma^2\right)\left(\sigma^2 - \hat{\sigma}^2\right) \tag{45}$$

So, After m^{th} iteration, the Fisher score algorithm is:

$$\sigma^2_{(m+1)} = \sigma^2_{(m)} + \mathcal{I}_m^{-1}\left(\sigma^2\right)\triangledown\left(\sigma^2_{(m)}\right) \tag{46}$$

Under the converge restriction which depends on the special requirements of real experiments, $\sigma^2_{(m+1)} \to \hat{\sigma}^2$ when $\mathcal{S}\left(\sigma^2_{(m+1)}\right) \approx 0$.

2.5 Measure of Imbalance

Since the number of observations of each level for random effects are different in unbalanced data, a measure is needed to test the imbalanceness of the data. Applied to model (1), the observation number n is also defined as the structure of observations in different levels of random effects.

$n = (n_1, n_2, \ldots n_m)$ and $\bar{n} = \frac{1}{m}\sum n_i$ where $m = p$ or q and $i = 1, \ldots, p$ or $1, \ldots, q$

There are three principles satisfied to construct the measures which are introduced by Ahrens and Pincus (1981). For example, a simple function of the 's symmetric in its arguments and reflect in a specified way properties of statistical analyses. The paper also proposed several principles satisfied measures as the candidates. These measures indentify to each other under some transformations. So, one of them applied in the article is given.

$$\nu_m(n) = \frac{1}{m\sum\left(\frac{n_i}{n}\right)^2} \tag{47}$$

where $n = \sum n_i$, $m = p \ or \ q$ and $i = 1, \ldots p$ or $1, \ldots q$.

We have $\frac{1}{m} \leq \nu_m(n) < 1$ in the unbalanced data and the smaller value denotes more imbalance. Largest $\nu_m(n) = 1$ is only for balanced data. Khuri, Mathew and Sinha (1998) showed that the sample variance of increases as the imbalance increasing.

For a two-way mixed model (1), $\nu_p(n)$ and $\nu_q(n)$ denote the imbalance for design matrix Z_1 and Z_2 and respectively. Here we suggest that the equation $\nu(n) = 0.5\nu_p(n) + 0.5\nu_q(n)$ is used to calculate the whole imbalance of the examples used in our essay.

3 Monte Carlo Comparison and Simulations

In order to compare variance components estimators from balanced to unbalanced data, the comparisons need to process under a variety of examples and true values of components. Swallow and Monahan (1984) illustrated that given the true values of variance components, the subgroup means and subgroup sums of squares are sufficient for the variance components estimators. This is exploited in our Monte Carlo simulation by using modified polar method (Marsglia and Bray, 1964) for generating normal random variables. The examples used to study the evaluations of modified Henderson's method 3 are given in APPENDIX A and are the same as in Al-sarraj and Rosen (2007). The reasons and questions about the examples choosing are discussed in section 6. The measure described in subsection 2.5 for test imbalance is utilized to show the difference of examples in subsection 3.1. The MSE effects of σ_2^2 to the σ_1^2 estimation of Henderson's method 3 and its modified are described in subsection 3.2. From the Table 3-1 and Table 3-2, the 'small' values ranges of σ_2^2 for different examples are obtained. The ranges of σ_2^2 suggest us to prefer to $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$ in terms of MSE based on comparison with $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$. The reason of using the 'small' values range of σ_2^2 is given in subsection 2.2 and 2.3. Then, from MSE effects of σ_2^2 and σ_1^2, we suggest a range $\sigma_2^2 < 0.1$ when $\sigma_1^2 = 0.1$ to apply all the examples. In this case, $\hat{\sigma}_{u1ML}^2$ and $\hat{\sigma}_{u1REML}^2$ are added to compare with four estimators of Henderson's method 3 and its modified. Hence, the bias and probability of getting negative estimate are used as the criteria to show the performances of six estimators. Furthermore, with the aim of extending our analysis to wide application, the range of ratio $\sigma_2^2/\sigma_1^2 < 1.0$ is checked. Since all the estimators should benefit from larger n, the difference of relationship between n and estimators are figured out in subsection 3.5.

3.1 Effects of Imbalance

The values of imbalance to show the differences between examples are given in table 3-1.

Table 3-1: The imbalace measure for each example

Example	n	p	q	$\nu_p(n)$	$\nu_q(n)$	$\nu(n)$
1	8	2	2	1	1	1
2	8	2	2	0.9412	0.9412	0.9412
3	8	2	2	0.8000	0.9412	0.8706
4	21	3	3	0.9439	0.9866	0.9653
5	30	3	3	0.8571	0.7937	0.8254
6	30	4	3	0.8858	0.8772	0.8815

Example 1 is balanced data, 2 and 4 are almost balanced. The examples 3, 5 and 6 are more unbalancedness than the others. In order to describe the relationship between the imbalance and the MSE of $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$. The observation n, p, q must be fixed. Since all the examples 1, 2 and 3 have $n = 8$, $p = 2$, $q = 2$, then this three examples are applied.

Hence, the true values of variance components $\sigma^2 = (1,1,1)$. The $MSE\left(\hat{\sigma}_{u1}^2\right)$ and $MSE\left(\hat{\sigma}_{11}^2\right)$ are calculated by equations (6) and (18).

Figure 3-1 clearly shows that $MSE\left(\hat{\sigma}_{u1}^2\right)$ are sensitive to the changing imbalance and have a increasing trend as the data becoming more imbalance. While $MSE\left(\hat{\sigma}_{11}^2\right)$ are similar with each other and also have a slight rising trend for larger imbalance. That means $\hat{\sigma}_{11}^2$ is more robust and performing better than $\hat{\sigma}_{u1}^2$ as the changes of imbalance.

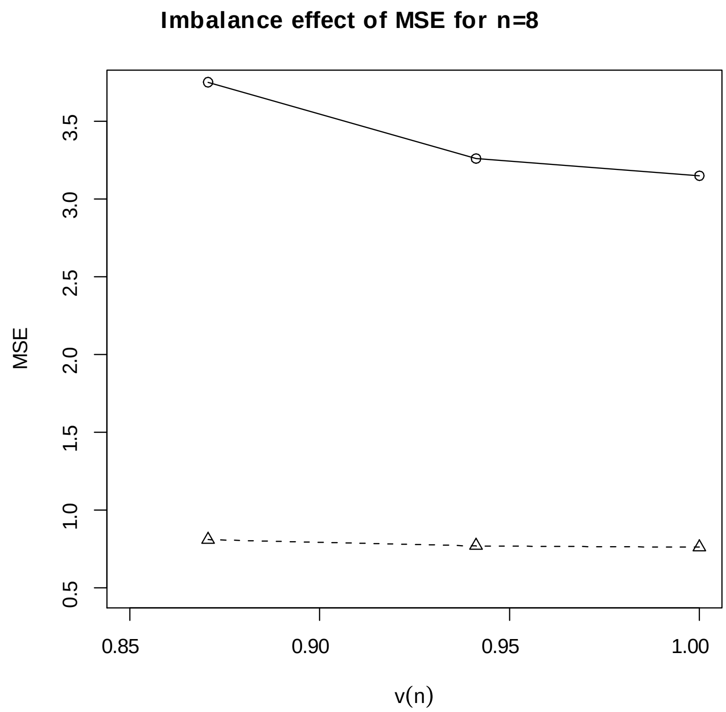

Figure 3-1: Imbalace effect of MSE for $n = 8, p = 2$ and $q = 2$. $\sigma_1^2 = 1, \sigma_2^2 = 1$ *and* $\sigma_e^2 = 1$. Solid line with circles is $MSE\left(\hat{\sigma}_{u1}^2\right)$.and the dashed line with triangles is $MSE\left(\hat{\sigma}_{11}^2\right)$

3.2 MSE Effets of σ_2^2

There are two Partitions for Henderson's method 3 and its modified. Based on the comparison between equations (6) and (11), equations (17) and (27), there exist a range of σ_2^2 to make $MSE\left(\hat{\sigma}_{u1}^2\right) < MSE\left(\hat{\sigma}_1^2\right)$ and $MSE\left(\hat{\sigma}_{11}^2\right) < MSE\left(\hat{\sigma}_{12}^2\right)$ .Then the main task of this part is to find the 'small' values range of σ_2^2 so that $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$ are recommended in terms of MSE compared with $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$ respectively. The true values used in our simulations are $\mu = 0$, $\sigma_1^2 = 0.1$, $\sigma_e^2 = 0.9$ and 10 different of σ_2^2 =0.01, 0.05, 0.1, 0.15, 0.25, 0.5, 0.75, 1, 1.5, 2 which range form 0.01 to 2. The equations to estimate $\hat{\sigma}_{u1}^2$, $\hat{\sigma}_1^2$, $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$ respectively are (5), (10), (14) and (24) based on $N = 1000$ simulations. The estimated biases are the difference between mean of estimates and true value $\sigma_1^2 = 0.1$. The observed MSE is calculated by the observed sample variance and estimated squared biases. The formula of observed MSE, estimated biases and sample mean are given in APPENDIX B. The observed MSE of $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_1^2$ to compare the predicted MSE in (6) and (11) are shown in Table 3-2. The 'small' values range of σ_2^2 to favor $\hat{\sigma}_{u1}^2$ are $\hat{\sigma}_1^2$ is also listed. Moreover, the observed MSE to compare with the predicted MSE of $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$ in (17) and (27) are given in Table 3-3 which is similar with Table 3-2.

Table 3-2: The observed MSE of $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_1^2$ for estimation of σ_1^2
based on 10 different σ_2^2, $\mu = 0, \sigma_1^2 = 0.1$ and $\sigma_e^2 = 0.9$ with $N = 1000$ simulations

		σ_2^2										
Ex.	Es.	0.01	0.05	0.1	0.15	0.25	0.5	0.75	1	1.5	2	'small' σ_2^2
1	$\hat{\sigma}_{u1}^2$	0.2351	0.2399	0.2453	0.2053	0.2129	0.2051	0.2417	0.2209	0.2340	0.2254	None
	$\hat{\sigma}_1^2$	0.2351	0.2399	0.2453	0.2053	0.2129	0.2051	0.2417	0.2209	0.2340	0.2254	
2	$\hat{\sigma}_{u1}^2$	0.2759	0.2420	0.2785	0.2575	0.2574	0.2429	0.2622	0.2462	0.2710	0.2693	σ_2^2<0.15
	$\hat{\sigma}_1^2$	0.2755	0.2440	0.2845	0.2604	0.2524	0.2423	0.2579	0.2463	0.2643	0.2583	
3	$\hat{\sigma}_{u1}^2$	0.4255	0.3340	0.3626	0.3556	0.4291	0.3418	0.3962	0.3467	0.4597	0.4496	σ_2^2<0.15
	$\hat{\sigma}_1^2$	0.4343	0.3418	0.3802	0.3552	0.4510	0.3359	0.3726	0.3159	0.3683	0.4073	
4	$\hat{\sigma}_{u1}^2$	0.0805	0.0918	0.1103	0.1246	0.1653	0.4000	0.5294	0.9613	1.4926	3.0523	σ_2^2<0.25
	$\hat{\sigma}_1^2$	0.1498	0.1876	0.1452	0.1364	0.1418	0.1480	0.1306	0.1536	0.1463	0.1424	
5	$\hat{\sigma}_{u1}^2$	0.0458	0.0447	0.0471	0.0499	0.0635	0.0952	0.1336	0.1977	0.3163	0.4955	σ_2^2<0.25
	$\hat{\sigma}_1^2$	0.0635	0.0568	0.0644	0.0541	0.0562	0.0603	0.0585	0.0587	0.0553	0.0589	
6	$\hat{\sigma}_{u1}^2$	0.0413	0.0560	0.0580	0.0673	0.1031	0.1839	0.2575	0.3835	0.5989	0.8223	σ_2^2<0.10
	$\hat{\sigma}_1^2$	0.0490	0.0575	0.0499	0.0530	0.0560	0.0508	0.0582	0.0527	0.0568	0.0563	

Table 3-3: The observed MSE of $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$ for estimation of σ_1^2
based on 10 different σ_2^2, $\mu = 0, \sigma_1^2 = 0.1$ and $\sigma_e^2 = 0.9$ with $N = 1000$ simulations

		σ_2^2										
Ex.	Es.	0.01	0.05	0.1	0.15	0.25	0.5	0.75	1	1.5	2	'small' σ_2^2
1	$\hat{\sigma}_{11}^2$	0.0269	0.0277	0.0280	0.0235	0.0249	0.0240	0.0270	0.0256	0.0266	0.0264	None
	$\hat{\sigma}_{12}^2$	0.0269	0.0277	0.0280	0.0235	0.0249	0.0240	0.0270	0.0256	0.0266	0.0264	
2	$\hat{\sigma}_{11}^2$	0.0312	0.0274	0.0312	0.0288	0.0292	0.0279	0.0292	0.0282	0.0305	0.0309	σ_2^2<0.25
	$\hat{\sigma}_{12}^2$	0.0311	0.0276	0.0318	0.0290	0.0287	0.0278	0.0287	0.0281	0.0297	0.0298	
3	$\hat{\sigma}_{11}^2$	0.0464	0.0363	0.0397	0.0392	0.0468	0.0376	0.0430	0.0383	0.0500	0.0495	σ_2^2<0.5
	$\hat{\sigma}_{12}^2$	0.0471	0.0370	0.0414	0.0391	0.0491	0.0369	0.0404	0.0352	0.0403	0.0445	
4	$\hat{\sigma}_{11}^2$	0.0182	0.0208	0.0236	0.0251	0.0308	0.0745	0.0954	0.1955	0.2844	0.5231	σ_2^2<0.25
	$\hat{\sigma}_{12}^2$	0.0277	0.0336	0.0275	0.0258	0.0261	0.0276	0.0246	0.0283	0.0268	0.0264	
5	$\hat{\sigma}_{11}^2$	0.0122	0.0120	0.0124	0.0127	0.0149	0.0196	0.0273	0.0340	0.0643	0.0914	σ_2^2<0.15
	$\hat{\sigma}_{12}^2$	0.0135	0.0126	0.0138	0.0125	0.0126	0.0133	0.0130	0.0125	0.0125	0.0127	
6	$\hat{\sigma}_{11}^2$	0.0137	0.0180	0.0176	0.0191	0.0299	0.0567	0.0821	0.1226	0.2198	0.3137	σ_2^2<0.10
	$\hat{\sigma}_{12}^2$	0.0170	0.0196	0.0176	0.0186	0.0189	0.0176	0.0197	0.0181	0.0193	0.0189	

From Table 3-2 and Table 3-3, the summaries we drawn are given below.

1. For the balanced data of example 1, the estimates of $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_1^2$ are equal to each other. The same situation is applied to $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$. In this case, the problem of lost unique estimator should not be considered.

2. The observed MSE of $\hat{\sigma}_{u1}^2$ are similar with $\hat{\sigma}_1^2$ in example 2 which is almost balanced data. In example 4, the MSE of $\hat{\sigma}_{u1}^2$ are smaller than the values in other examples when σ_2^2 is small. But it has terrible result if σ_2^2 is large. For the examples 3, 5 and 6, both $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$ have a gradually increasing trend as σ_2^2 increases. Since $MSE\left(\hat{\sigma}_1^2\right)$ and $MSE\left(\hat{\sigma}_{12}^2\right)$ do not depend on σ_2^2, their observed MSE stay stationary. The MSE of all four estimators benefit from the larger n.

3. For fixed $\sigma_1^2 = 0.1$ and changes σ_2^2, both $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$ have achieved substantially improvement compared with $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_1^2$ respectively in terms of MSE.

4. The 'small' values ranges of σ_2^2 are listed to prefer to $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$ compared with $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$ respectively. The upper bounds are around from 0.10 to 0.50. So the 'small' values range $\sigma_2^2 < 0.1$ is recommended for applied to all the examples except example 1.

3.3 MSE Effects of σ_1^2

$\sigma_2^2 < 0.1$ is recommended as the 'small' values range to favor $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$, based on the analysis in subsection 3.2. It is easy to see that, the MSE of Henderson's method 3 and its modified depend on σ_1^2. If we choose one value from $\sigma_2^2 < 0.1$, there should also have a range of σ_1^2 to favor $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$. In order to figure out the relationship between the estimators and σ_1^2 in this subsection, we give 10 different values of σ_1^2=0.001, 0.01, 0.05, 0.1, 0.15, 0.2, 0.5, 1, 2, 5 which range from 0.001 to 5. $\mu = 0$ and $\sigma_e^2 = 0.9$ are simulated. $\sigma_2^2 = 0.05$ is chosen from the 'small' values range. The simulation number is 1000. Commonly used methods $\hat{\sigma}_{u1ML}^2$ and $\hat{\sigma}_{u1REML}^2$ are considered to compare with the estimators of Henderson's method 3 and its modified. The $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$ are eliminated in example 1 because the balanced data has the same estimates for Partition I and II. The observed MSE, estimated biases applied are same as subsection 3.2. Then, the observed MSE for different estimators of σ_1^2 are given in Table 3-4. And the estimated biases for all the examples of different σ_1^2 are presented in Table 3-5.

From Table 3-4 and Table 3-5, the summaries we draw are given below.

1. The observed MSE of $\hat{\sigma}_{u1}^2$ are lower than $\hat{\sigma}_1^2$ except in the example 1 and 2. Example 1 is balanced data and example 2's imbalance is closed to 1. It is reasonable to see that the estimates are same in example 1 and similar with each other in example 2. This situation also applied to the MSE comparison between $\hat{\sigma}_{11}^2$, and $\hat{\sigma}_{12}^2$. So the condition of 'small' value given by $\sigma_2^2 = 0.05$ is sufficient to confirm us to choose $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_1^2$ rather than $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$. The results also show us that the modified estimator improves unmodified one in terms of MSE.
2. The MSE of $\hat{\sigma}_{u1ML}^2$ are smaller than $\hat{\sigma}_{u1REML}^2$ for each example, though it have serious bias if σ_1^2 is large. So, $\hat{\sigma}_{u1ML}^2$ performs better than $\hat{\sigma}_{u1REML}^2$ in terms of MSE. Moreover, the MSE of $\hat{\sigma}_{u1ML}^2$ are also approximate equal to $\hat{\sigma}_{11}^2$ and they have lower values than the others. Hence, $\hat{\sigma}_{u1ML}^2$ and $\hat{\sigma}_{11}^2$ can be recommended when MSE is concerned.
3. The biases of $\hat{\sigma}_{11}^2$, $\hat{\sigma}_{12}^2$ and $\hat{\sigma}_{u1ML}^2$ increase dramatically, and will have terrible results if σ_1^2 is large. Whereas, the unbiased estimators $\hat{\sigma}_{u1}^2$, $\hat{\sigma}_1^2$ and $\hat{\sigma}_{u1REML}^2$ are more robust and approximately equal to 0. Then $\hat{\sigma}_{u1REML}^2$ and Henderson's method 3 are recommended if the unbiasedness is the main concern.

Table 3-4: Observed MSE for estimators of σ_1^2 based on 10 different σ_1^2, $\mu = 0, \sigma_2^2 = 0.05$ and $\sigma_e^2 = 0.9$ with $N = 1000$ simulations

		σ_1^2									
E	Es.	0.001	0.01	0.05	0.1	0.15	0.2	0.5	1.0	2.0	5.0
1	$\hat{\sigma}^2_{u1}$	0.1238	0.1250	0.1719	0.2427	0.3205	0.3607	1.1263	3.1653	9.5649	56.0616
	$\hat{\sigma}^2_{11}$	0.0132	0.0128	0.0184	0.0282	0.0395	0.0512	0.2252	0.7693	2.8261	17.4888
	$\hat{\sigma}^2_{u1REML}$	0.0941	0.0890	0.1351	0.1990	0.2704	0.2997	1.0359	3.0396	9.4204	55.7817
	$\hat{\sigma}^2_{u1ML}$	0.0175	0.0171	0.0258	0.0404	0.0586	0.0731	0.3109	1.0109	3.4654	20.7649
2	$\hat{\sigma}^2_{u1}$	0.1658	0.1287	0.2179	0.2562	0.4086	0.4303	1.1886	3.2356	10.0367	50.6879
	$\hat{\sigma}^2_{1}$	0.1699	0.1316	0.2153	0.2568	0.4080	0.4292	1.1969	3.2513	10.0444	50.5883
	$\hat{\sigma}^2_{11}$	0.0177	0.0130	0.0229	0.0298	0.0483	0.0583	0.2227	0.7667	2.9140	16.9085
	$\hat{\sigma}^2_{12}$	0.0182	0.0133	0.0227	0.0298	0.0483	0.0581	0.2230	0.7680	2.9151	16.9052
	$\hat{\sigma}^2_{u1REML}$	0.1310	0.0869	0.1728	0.2017	0.3448	0.3707	1.1007	3.1020	9.8299	50.3896
	$\hat{\sigma}^2_{u1ML}$	0.0258	0.0153	0.0339	0.0420	0.0758	0.0882	0.3144	1.0235	3.6079	19.3853
3	$\hat{\sigma}^2_{u1}$	0.2285	0.1931	0.2795	0.3333	0.4257	0.5784	1.5486	3.1368	9.9593	48.1104
	$\hat{\sigma}^2_{1}$	0.2397	0.2119	0.2935	0.3481	0.4611	0.5929	1.5282	3.1542	10.2051	48.4232
	$\hat{\sigma}^2_{11}$	0.0243	0.0202	0.0292	0.0368	0.0512	0.0719	0.2515	0.7696	2.8263	16.7087
	$\hat{\sigma}^2_{12}$	0.0253	0.0220	0.0305	0.0381	0.0545	0.0728	0.2496	0.7686	2.8419	16.7021
	$\hat{\sigma}^2_{u1REML}$	0.1771	0.1476	0.2190	0.2598	0.3599	0.4822	1.4154	2.9606	9.8117	47.6020
	$\hat{\sigma}^2_{u1ML}$	0.0340	0.0247	0.0381	0.0493	0.0739	0.1044	0.3842	1.0216	3.5289	18.9133
4	$\hat{\sigma}^2_{u1}$	0.0478	0.0465	0.0702	0.0747	0.1059	0.1335	0.4492	1.3921	4.6656	27.0626
	$\hat{\sigma}^2_{1}$	0.0710	0.0814	0.1140	0.1372	0.1829	0.2516	0.6670	2.1822	6.9807	34.7800
	$\hat{\sigma}^2_{11}$	0.0077	0.0078	0.0137	0.0165	0.0276	0.0380	0.1646	0.5949	2.1830	13.1927
	$\hat{\sigma}^2_{12}$	0.0122	0.0139	0.0197	0.0258	0.0390	0.0543	0.1997	0.7115	2.5678	14.8997
	$\hat{\sigma}^2_{u1REML}$	0.0177	0.0227	0.0460	0.0592	0.0915	0.1108	0.4229	1.5222	4.5663	25.9121
	$\hat{\sigma}^2_{u1ML}$	0.0050	0.0075	0.0156	0.0216	0.0400	0.0501	0.2177	0.8233	2.5913	14.8623
5	$\hat{\sigma}^2_{u1}$	0.0149	0.0156	0.0256	0.0484	0.0631	0.0957	0.4381	1.3197	4.4046	28.6096
	$\hat{\sigma}^2_{1}$	0.0174	0.0191	0.0354	0.0578	0.0775	0.1228	0.5881	1.6704	5.4763	38.2848
	$\hat{\sigma}^2_{11}$	0.0032	0.0033	0.0061	0.0129	0.0204	0.0302	0.1665	0.5750	2.1864	13.3664
	$\hat{\sigma}^2_{12}$	0.0029	0.0033	0.0067	0.0131	0.0213	0.0331	0.1857	0.6303	2.4277	15.1549
	$\hat{\sigma}^2_{u1REML}$	0.0118	0.0129	0.0189	0.0434	0.0578	0.0988	0.3894	1.2360	4.1832	27.1272
	$\hat{\sigma}^2_{u1ML}$	0.0035	0.0042	0.0068	0.0175	0.0273	0.0454	0.2088	0.6882	2.4131	14.8074
6	$\hat{\sigma}^2_{u1}$	0.0248	0.0284	0.0377	0.0529	0.0721	0.0956	0.3129	0.9883	3.3501	20.3179
	$\hat{\sigma}^2_{1}$	0.0272	0.0253	0.0380	0.0583	0.0758	0.1070	0.3227	1.0370	3.4599	21.6717
	$\hat{\sigma}^2_{11}$	0.0061	0.0073	0.0101	0.0166	0.0245	0.0350	0.1407	0.4907	1.8164	10.7648
	$\hat{\sigma}^2_{12}$	0.0087	0.0080	0.0122	0.0198	0.0276	0.0399	0.1482	0.5039	1.8648	11.0945
	$\hat{\sigma}^2_{u1REML}$	0.0119	0.0117	0.0203	0.0392	0.0514	0.0735	0.2815	0.9111	3.1506	18.7353
	$\hat{\sigma}^2_{u1ML}$	0.0048	0.0048	0.0090	0.0196	0.0288	0.0429	0.1781	0.5855	2.0512	11.9087

Table 3-5: Estimated Biases for estimators of σ_1^2 based on 10 different σ_1^2, $\mu = 0$, $\sigma_2^2 = 0.05$ and $\sigma_e^2 = 0.9$ with $N = 1000$ simulations

		σ_1^2									
E	Es.	0.001	0.01	0.05	0.1	0.15	0.2	0.5	1.0	2.0	5.0
1	$\hat{\sigma}_{u1}^2$	0.0168	-0.0105	-0.0133	-0.0320	-0.0031	0.0369	-0.0012	-0.0141	-0.0702	-0.2245
	$\hat{\sigma}_{11}^2$	0.0257	0.0112	-0.0164	-0.0553	-0.0796	-0.1004	-0.3125	-0.6499	-1.3351	-3.3860
	$\hat{\sigma}_{u1REML}^2$	0.1290	0.1109	0.1021	0.0796	0.1013	0.1300	0.0814	0.0541	-0.0115	-0.1802
	$\hat{\sigma}_{u1ML}^2$	0.0438	0.0317	0.0043	-0.0328	-0.0505	-0.0630	-0.2467	-0.5142	-1.0530	-2.6414
2	$\hat{\sigma}_{u1}^2$	0.0047	-0.0208	0.0178	0.0050	0.0396	-0.0283	-0.0110	-0.0100	0.0078	0.3624
	$\hat{\sigma}_1^2$	0.0056	-0.0220	0.0188	0.0051	0.0421	-0.0278	-0.0119	-0.0059	0.0029	0.3662
	$\hat{\sigma}_{11}^2$	0.0233	0.0090	-0.0047	-0.0420	-0.0643	-0.1198	-0.3140	-0.6476	-1.3079	-3.1896
	$\hat{\sigma}_{12}^2$	0.0237	0.0086	-0.0043	-0.0419	-0.0634	-0.1196	-0.3142	-0.6462	-1.3094	-3.1882
	$\hat{\sigma}_{u1REML}^2$	0.1389	0.1087	0.1411	0.1243	0.1452	0.0820	0.0746	0.0650	0.0622	0.4129
	$\hat{\sigma}_{u1ML}^2$	0.0493	0.0284	0.0194	-0.0156	-0.0325	-0.0901	-0.2573	-0.5117	-1.0236	-2.3457
3	$\hat{\sigma}_{u1}^2$	-0.0087	-0.0282	0.0441	-0.0342	-0.0017	0.0169	0.0288	0.0139	-0.0226	-0.0169
	$\hat{\sigma}_1^2$	-0.0035	-0.0284	0.0390	-0.0355	-0.0011	0.0187	0.0308	0.0217	-0.0264	-0.0142
	$\hat{\sigma}_{11}^2$	0.0254	0.0131	0.0095	-0.0498	-0.0718	-0.0988	-0.2950	-0.6332	-1.3112	-3.3106
	$\hat{\sigma}_{12}^2$	0.0275	0.0136	0.0083	-0.0496	-0.0710	-0.0977	-0.2939	-0.6301	-1.3119	-3.3091
	$\hat{\sigma}_{u1REML}^2$	0.1602	0.1453	0.1875	0.1180	0.1359	0.1597	0.1326	0.1177	0.0608	0.0448
	$\hat{\sigma}_{u1ML}^2$	0.0468	0.0374	0.0313	-0.0301	-0.0569	-0.0705	-0.2496	-0.5153	-1.0501	-2.5660
4	$\hat{\sigma}_{u1}^2$	-0.0025	0.0035	-0.0071	-0.0068	-0.0093	0.0012	0.0012	0.0020	0.0003	0.0930
	$\hat{\sigma}_1^2$	-0.0162	0.0068	-0.0015	-0.0054	-0.0320	-0.0036	0.0108	-0.0004	0.0318	0.1519
	$\hat{\sigma}_{11}^2$	0.0098	0.0097	-0.0185	-0.0417	-0.0687	-0.0872	-0.2409	-0.4966	-1.0026	-2.4809
	$\hat{\sigma}_{12}^2$	0.0047	0.0088	-0.0176	-0.0486	-0.0886	-0.1058	-0.2750	-0.5710	-1.1404	-2.8384
	$\hat{\sigma}_{u1REML}^2$	0.0497	0.0539	0.0329	0.0287	0.0060	0.0214	0.0032	-0.0219	-0.0377	0.0552
	$\hat{\sigma}_{u1ML}^2$	0.0218	0.0189	-0.0102	-0.0324	-0.0645	-0.0751	-0.1946	-0.3794	-0.7295	-1.6692
5	$\hat{\sigma}_{u1}^2$	-0.0007	-0.0047	-0.0017	-0.0058	0.0091	0.0165	-0.0237	-0.0215	-0.0087	-0.0984
	$\hat{\sigma}_1^2$	-0.0021	-0.0068	-0.0011	-0.0023	0.0061	0.0139	-0.0447	-0.0123	0.0010	-0.1157
	$\hat{\sigma}_{11}^2$	0.0045	-0.0016	-0.0212	-0.0499	-0.0691	-0.0912	-0.2680	-0.5301	-1.0496	-2.6685
	$\hat{\sigma}_{12}^2$	0.0025	-0.0048	-0.0259	-0.0557	-0.0816	-0.1077	-0.3079	-0.5878	-1.1688	-2.9767
	$\hat{\sigma}_{u1REML}^2$	0.0375	0.0319	0.0296	0.0187	0.0246	0.0351	-0.0174	-0.0175	0.0125	-0.1187
	$\hat{\sigma}_{u1ML}^2$	0.0163	0.0075	-0.0121	-0.0367	-0.0511	-0.0615	-0.2017	-0.3667	-0.6881	-1.7696
6	$\hat{\sigma}_{u1}^2$	-0.0001	0.0020	-0.0049	0.0003	0.0137	-0.0071	0.0022	-0.0481	0.0465	0.0970
	$\hat{\sigma}_1^2$	-0.0048	-0.0025	0.0005	-0.0044	0.0127	-0.0031	0.0165	-0.0466	0.0319	0.0711
	$\hat{\sigma}_{11}^2$	0.0088	0.0065	-0.0130	-0.0326	-0.0440	-0.0775	-0.1970	-0.4318	-0.7952	-2.0115
	$\hat{\sigma}_{12}^2$	0.0047	0.0021	-0.0137	-0.0381	-0.0501	-0.0808	-0.2000	-0.4530	-0.8428	-2.1236
	$\hat{\sigma}_{u1REML}^2$	0.0422	0.0407	0.0257	0.0170	0.0311	0.0113	0.0004	-0.0430	0.0332	0.0614
	$\hat{\sigma}_{u1ML}^2$	0.0223	0.0181	-0.0056	-0.0271	-0.0322	-0.0616	-0.1500	-0.3051	-0.4986	-1.2226

Probability of Getting Negative Estimate As a limitation for Henderson's method 3, there exist negative estimates. The formula of observed probability of getting negative estimate is given in APPENDIX C. Since the iterative algorithms are used to estimate $\hat{\sigma}_{u1ML}^2$ and $\hat{\sigma}_{u1REML}^2$, the negative estimate condition must be taken into account in the computer programs for solving their equations; see Searle, Casella and McCulloch (1992). The probability of getting negative estimate by ML and REML are equal to 0 and need not to be considered. The reason of eliminating $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$ in the example 1 is that Henderson's method 3 and its modified do not have the problem of lost unique estimators. The observed probability of getting negative estimate of $\hat{\sigma}_{u1}^2$, $\hat{\sigma}_1^2$, $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$ are listed in Table 3-6.

Table 3-6: The observed Probability of getting negative estimate for estimation of σ_1^2 based on 10 different σ_1^2, μ=0, σ_2^2=0.05 and σ_e^2=0.9 with $N = 1000$ simulations

		σ_1^2									
Ex.	Es.	0.001	0.01	0.05	0.1	0.15	0.2	0.5	1.0	2.0	5.0
1	$\hat{\sigma}_{u1}^2$	0.646	0.643	0.580	0.569	0.511	0.489	0.423	0.324	0.218	0.169
	$\hat{\sigma}_{11}^2$	0.569	0.567	0.519	0.487	0.440	0.428	0.364	0.269	0.193	0.151
2	$\hat{\sigma}_{u1}^2$	0.646	0.637	0.582	0.580	0.524	0.488	0.392	0.323	0.251	0.158
	$\hat{\sigma}_1^2$	0.635	0.645	0.590	0.574	0.531	0.491	0.396	0.324	0.247	0.159
	$\hat{\sigma}_{11}^2$	0.569	0.575	0.512	0.507	0.466	0.429	0.343	0.279	0.207	0.140
	$\hat{\sigma}_{12}^2$	0.569	0.568	0.517	0.500	0.460	0.435	0.346	0.285	0.207	0.139
3	$\hat{\sigma}_{u1}^2$	0.640	0.619	0.600	0.592	0.545	0.522	0.404	0.361	0.275	0.188
	$\hat{\sigma}_1^2$	0.641	0.626	0.603	0.586	0.545	0.530	0.399	0.353	0.272	0.195
	$\hat{\sigma}_{11}^2$	0.556	0.542	0.516	0.523	0.488	0.461	0.354	0.304	0.224	0.161
	$\hat{\sigma}_{12}^2$	0.566	0.549	0.539	0.515	0.480	0.458	0.339	0.309	0.237	0.166
4	$\hat{\sigma}_{u1}^2$	0.539	0.515	0.433	0.385	0.365	0.323	0.202	0.127	0.064	0.024
	$\hat{\sigma}_1^2$	0.602	0.579	0.545	0.499	0.476	0.421	0.300	0.220	0.130	0.065
	$\hat{\sigma}_{11}^2$	0.541	0.521	0.422	0.374	0.344	0.298	0.186	0.121	0.059	0.025
	$\hat{\sigma}_{12}^2$	0.564	0.539	0.501	0.470	0.447	0.386	0.279	0.201	0.120	0.058
5	$\hat{\sigma}_{u1}^2$	0.602	0.586	0.469	0.430	0.363	0.266	0.184	0.099	0.047	0.016
	$\hat{\sigma}_1^2$	0.644	0.609	0.508	0.450	0.392	0.320	0.234	0.140	0.062	0.019
	$\hat{\sigma}_{11}^2$	0.570	0.545	0.446	0.394	0.340	0.240	0.173	0.089	0.038	0.011
	$\hat{\sigma}_{12}^2$	0.618	0.578	0.477	0.425	0.370	0.301	0.215	0.127	0.056	0.018
6	$\hat{\sigma}_{u1}^2$	0.538	0.497	0.425	0.376	0.316	0.233	0.121	0.066	0.022	0.006
	$\hat{\sigma}_1^2$	0.603	0.569	0.476	0.407	0.341	0.288	0.159	0.080	0.034	0.009
	$\hat{\sigma}_{11}^2$	0.537	0.502	0.398	0.337	0.294	0.212	0.102	0.052	0.018	0.005
	$\hat{\sigma}_{12}^2$	0.570	0.537	0.446	0.370	0.313	0.261	0.146	0.073	0.033	0.007

Results in table 3-6 show that the values of probability of getting negative estimate of $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$ are similar with each other as well as $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$ in different examples. It is reasonable to have that the negative probability two Partitions of Henderson's method 3 and its modified decrease for larger σ_1^2. The modified estimators $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$ have smaller values than unmodified ones. That means modified estimator perform better than unmodified one when the negative probability is concerned.

3.4 The Ratio σ_2^2/σ_1^2 Test

The 'small' values range of $\sigma_2^2 < 0.1$ is obtained from the MSE comparison in subsection 3.2 and 3.3. Generally, the true values of variance components are varied for a large range. Here we need to extend this 'small' values range to the ratio σ_2^2/σ_1^2 with the aim of wide application. The range of ratio $\sigma_2^2/\sigma_1^2 < 1.0$ should be recommended based on the calculation from $\sigma_2^2 < 0.1$ and $\sigma_1^2 = 0.1$ in subsection 3.2. Let us choose one value $\sigma_2^2/\sigma_1^2 = 0.8$ in the ratio range. And for the same ratio, there exist different values of σ_2^2 and σ_1^2. Here we give the true values for simulation σ_2^2=0.8, 4, 12, 24, 40, 80 and σ_1^2=1, 5, 15, 50, 100. Hence, the range of σ_2^2 and σ_1^2 could cover many true values of variance components in real experiments. The other parameter is $\mu = 0$ and $\sigma_e^2 = 0.9$. Examples 2 and 5 are used to test the ratio based on $N = 1000$ simulations. The observed MSE of Henderson's method 3 and its modified are given in Table 3-7.

Table 3-7: The observed MSE for estimation of σ_1^2 based on $\sigma_2^2/\sigma_1^2 = 0.8$, $\mu = 0$, $\sigma_e^2 = 0.9$ with $N = 1000$ simulations

		$\sigma_2^2/\sigma_1^2 = 0.8$					
Ex.	Es	0.8, 1	4, 5	12, 15	24, 30	40, 50	80, 100
	$\hat{\sigma}_{u1}^2$	2.8979	57.3405	423.4799	995.3098	4847.170	19240.168
2	$\hat{\sigma}_1^2$	2.9417	56.0465	420.1413	1015.9780	4848.737	19415.408
	$\hat{\sigma}_{11}^2$	0.7336	17.3185	145.6219	383.3311	1657.436	6583.363
	$\hat{\sigma}_{12}^2$	0.7399	17.2168	145.4236	384.9611	1653.197	6612.680
	$\hat{\sigma}_{u1}^2$	1.7133	34.8602	267.2678	1204.2242	3348.202	12375.354
5	$\hat{\sigma}_1^2$	1.7311	36.6264	274.2066	1221.0254	3351.043	13188.749
	$\hat{\sigma}_{11}^2$	0.6093	13.5307	113.3753	475.6935	1356.286	5197.689
	$\hat{\sigma}_{12}^2$	0.6408	14.9743	125.7473	520.3050	1447.766	5734.068

Table 3-7 shows us that for example 2, the MSE of $\hat{\sigma}_{u1}^2$ are larger than $\hat{\sigma}_1^2$ sometimes as well as the comparison between $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$. Then, the ratio range is not stable when n is low which will be discussed in section 6. Moreover, the observed MSE of $\hat{\sigma}_{u1}^2$ are smaller than $\hat{\sigma}_1^2$ with ratio $\sigma_2^2/\sigma_1^2 = 0.8$ in example 5. The MSE values of $\hat{\sigma}_{11}^2$ are also lower than $\hat{\sigma}_1^2$. That means if n is large the 'small' range values of can be extended to the ratio $\sigma_2^2/\sigma_1^2 < 1.0$. If the MSE is the main interest, the modified estimator is favored $\hat{\sigma}_{11}^2$ over the other three estimators in this case.

3.5 MSE Effects of n

As described in subsection 3.2 and 3.3, all the estimators benefit from the larger n. The main task of this part is to figure out effects of n to the MSE of different estimators. $\sigma_1^2 = 1$, $\sigma_2^2 = 0.05$ and $\sigma_e^2 = 0.9$, $\mu = 0$, are used as simulated values. The MSE of $\hat{\sigma}_{u1}^2$, $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{u1ML}^2$ are calculated here. The other three estimators are eliminated because of the MSE comparison in subsection 3.3. Example 5 is chosen as the basic experiment. 4 different observations n=30, 150, 450, 900 are applied here based on $N = 1000$ simulations. The observed MSE results with different n are drawn in Figure 3-2.

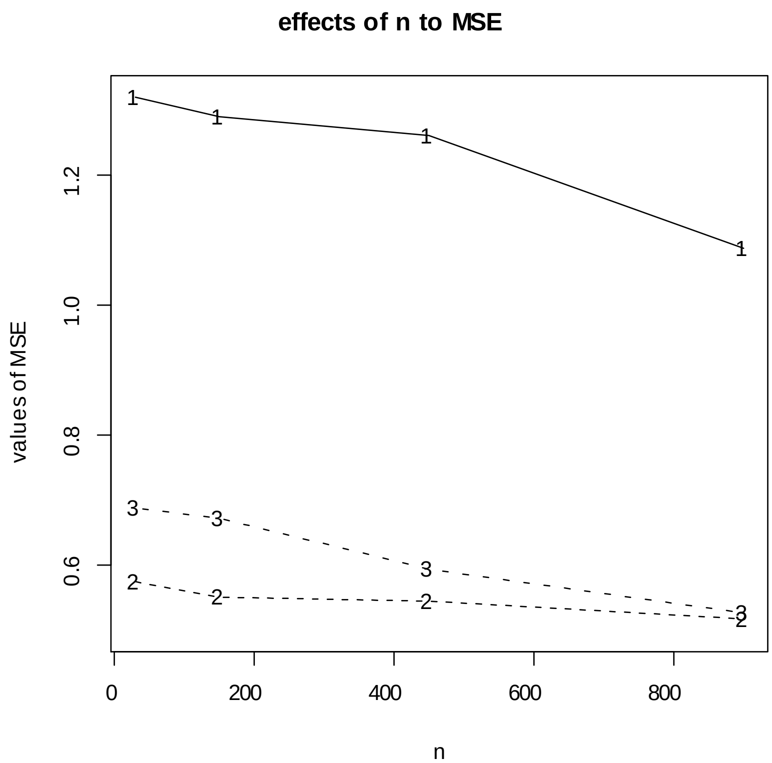

Figure 3-2: Observed MSE for different n=30, 150, 450, 900 based on σ_2^2=0.05, σ_1^2=1,σ_e^2=0.9 and μ=0 with N=1000 simulations.
The 1 with solid line is $\hat{\sigma}_{u1}^2$, the 2 with dashed line is $\hat{\sigma}_{11}^2$ and 3 is $\hat{\sigma}_{u1ML}^2$

Figure 3-2 shows us that the gaps between different estimators become smaller as the increases of n. The MSE of $\hat{\sigma}_{u1}^2$ are more sensitive for the changing of n than $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{u1ML}^2$. If n is large enough, then the MSE of the estimators are approximately equal to each other. So in this case, the unbiased estimators are preferred to the biased ones.

Conclusion 4 *$\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{11}^2$ are preferred to $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$ respectively in terms of MSE, based on the condition that 'small' values range of $\sigma_2^2 < 0.1$. The imbalance of data and n also have effects to the range of 'small' values. We also extend the 'small' values of σ_2^2 to the ratio $\sigma_2^2/\sigma_1^2 < 1.0$ with the aim of wide application. Although the modified $\hat{\sigma}_{11}^2$ do have big biases if σ_2^2 is large, its MSE is smaller than other estimators for all the examples. If σ_2^2 is larger enough and exceeds the boudary value, then $\hat{\sigma}_{11}^2$ can be replaced by $\hat{\sigma}_{12}^2$. Moreover, since its unbiasedness, $\hat{\sigma}_{u1REML}^2$ is more robust than $\hat{\sigma}_{u1ML}^2$. But there is no reason to choose $\hat{\sigma}_{u1REML}^2$ in terms of MSE. This same result can also be found in Kelly and Mathew (1994). $\hat{\sigma}_{u1ML}^2$ and $\hat{\sigma}_{11}^2$ both are recommended for small when MSE is concerned. However, because of its noniterative calculation, $\hat{\sigma}_{11}^2$ is chosen as an appropriate estimator compared with $\hat{\sigma}_{u1ML}^2$.*

4 Split-Plot Design Experiment Application

In this section, the commonly used mixed model of split-plot design is implemented with the estimators in our previous discussion. Ramon (1996) introduce the split-plot design theory. The split-plot design involves often two experimental factors, A and B which are divided into the main plots and subplots respectively. Levels of A are randomly assigned to whole plots, and levels of B are randomly assigned to split plots within each whole plot. The design provides more precise information about B than about A, and it often arises when A can be applied only to large experiments units. An example from Ramon (1996) is where A represents irrigation levels for large plots of land and B represents different crop varieties planted in each large plot.

4.1 Data Description

The data given in Ramon (1996) is obtained from a balanced split-plot design with the whole plots arranged in a randomized complete-block design. The whole-plot factor is denoted by A, and the subplot factor is B. The A, B and Block are classification variables. Table 4-1 gives the data for application.

Figure 4-1: Split-Plot design data from Ramon(1996)

No.	Block	A	B	Y	No.	Block	A	B	Y
1	1	1	1	56	13	1	2	1	41
2	1	2	1	50	14	1	2	2	36
3	1	3	1	39	15	1	2	3	35
4	2	1	1	30	16	2	2	1	25
5	2	2	1	36	17	2	2	2	28
6	2	3	1	33	18	2	2	3	30
7	3	1	1	32	19	3	2	1	24
8	3	2	1	31	20	3	2	2	27
9	3	3	1	15	21	3	2	3	19
10	4	1	1	30	22	4	2	1	25
11	4	2	1	35	23	4	2	2	30
12	4	3	1	17	24	4	2	3	18

The observation $n = 24$. There are 3 levels in A, 2 levels in B and 4 levels in Block. Let $\bar{Y} = 30.9167$ denote the mean of response Y, and $D(Y) = 93.5580$ denote sample variance. The distribution of response is drawn as Figure 4-1.

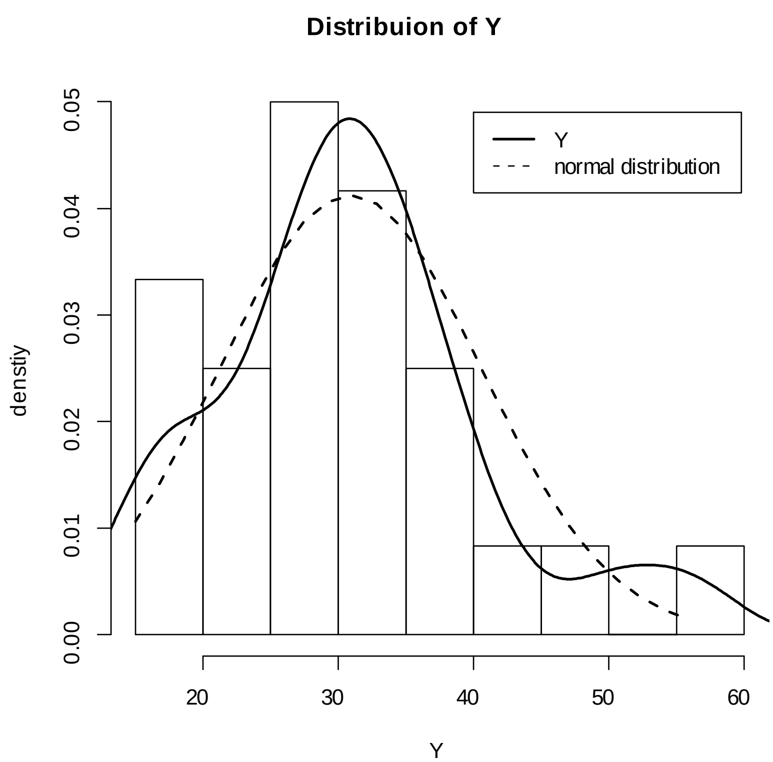

Figure 4-1: The solid line denotes the distribution of response Y, the dashed line is normal distribution with $N(\bar{Y}, D(Y))$

Figure 4-1 shows us that the distribution of response vector Y are approximately to normal distribution with $MVN(\bar{Y}, D(Y))$. That satisfys the assumptions of model (1) given in subsection 2.1.

4.2 Modelling and Application

Ramon (1996) suggested us to construct a two-way mixed model which is the same as model (1). The variables of A, B and A*B are seen as fixed effects, A*Block and Block are seen as random effects. Since A*Block has both the effects of A and Block, and the levels of A are randomly assigned to the main plots ,here the variance component of A*Block is the same as σ_1^2. Let u_1 and u_2 denote A*Block and Block respectively. The model (1) applied to data in Table 4-1 is described below.

The design matrix for fixed effects contains 24 columns and 6 rows. Here, $p = 12$ and $q = 4$ are the levels of random effects u_1 and u_2 which are distributed as $MVN\left(0, \sigma_1^2 I_{12}\right)$, $MVN\left(0, \sigma_2^2 I_4\right)$ based on the assumption. Then the design matrix for random effects, $Z_{1(24\times12)}$ and $Z_{1(24\times4)}$ is obtained from the data[6].

The results of $\hat{\sigma}^2_{u1}$, $\hat{\sigma}^2_{11}$, $\hat{\sigma}^2_{u1ML}$ and $\hat{\sigma}^2_{u1REML}$ in this case are given in Table 4-2.

Table 4-2:The estimates results with split-plots design application

	$\hat{\sigma}^2_{u1}$	$\hat{\sigma}^2_{11}$	$\hat{\sigma}^2_{u1ML}$	$\hat{\sigma}^2_{u1REML}$
estimate	-155695.5	37.15833	8.946263	15.3819

From Table 4-2, we can see that the estimate of $\hat{\sigma}^2_{u1}$ is negative and would cause terrible problem. The value of $\hat{\sigma}^2_{u1ML}$ is not close to $\hat{\sigma}^2_{u1REML}$. $\hat{\sigma}^2_{11}$ has achieved improvement for $\hat{\sigma}^2_{u1}$.

5 Conclusion

The aim of our article is to evaluate the performance of modified Henderson's method 3 developed by Al-Sarraj and Rosen (2007) by means of simulations. The model we used is a two-way linear mixed model satisfying several assumptions. Six examples from unbalanced to balanced data are considered. Several criteria MSE, bias and probability of getting negative estimate are used to show the performance of the modified estimator together with the unmodified one, ML and REML.

For the unbalanced data, in order to solve the problem of no unique estimators, the estimation of σ_1^2 by Henderson's method 3 are divided into Partition I and Partition II. The modified Henderson's method 3 is also divided into two Partitions. Hence we choose one of the Partitions with smaller MSE as an appropriate estimator. From the simulation results in section 3, the two Partitions are equal to each other. The same results also exist in the modified estimators. A 'small' values range of $\sigma_2^2 < 0.1$ is obtained from the MSE comparison between the two Partitions. Then, we recommend a ratio range of $\sigma_2^2/\sigma_1^2 < 1.0$ to decide which Partition is preferred. This ratio range can apply to both modified and unmodified Henderson' method 3 when n is large. Hence, The modified estimator have achieved great improvement compared with unmodified one in terms of MSE. Moreover, If the negative probability is concerned, then the modified estimator also performs better than unmodified one. However, if bias is considered, the modified Henderson's method 3 and $\hat{\sigma}^2_{u1ML}$ is not suggested. The MSE of is close to the mofidified estimator sometimes. $\hat{\sigma}^2_{u1ML}$ can also recommended if the MSE is concerned.

From the simulation analysis and split-plot design experiment application, we have the conclusion that the modified estimators can be considered as an appropriate estimator if the MSE and probability of getting negative estimate are concerned. Furthermore, with the explicit expression and noniterative computation, modified estimator is also computationally faster than ML and REML and performs better in terms of MSE sometimes.

6 Discussion

As Swallow and Searle (1978) illustrated, we can not choose a set of examples which could cover all possible unbalancedness. The usual mind to select examples is from slightly unbalanced to badly unbalanced. Here, the motivation of Al-Sarraj and Rosen (2007) to select the examples is to compare the MSE of Partition I and Partition

[6]The details to obtain the design matrixes of X, $Z_{1(24\times12)}$ and $Z_{1(24\times4)}$ are given in APPENDIX C

II of Henderson's method 3. All the examples applied are not existing terrible unbalancedness The observation n for examples 1 to 3 is 8 which can be seen as low number of observations. We have shown that the theoretical relationship between σ_2^2 and the MSE of Partition I. The MSE of Partition I should have a increasing trend for larger σ_2^2 if σ_1^2 and σ_e^2 are fixed. From the simulation results of the examples with $n = 8$ in Table 3-2 and Table 3-3, the MSE of Partition I do not have a obviously ascending trend as σ_2^2 increasing. However, if $n \geq 21$, then this increasing trend presents obviously. The ratio $\sigma_2^2/\sigma_1^2 < 1.0$ recommended for wide application also perform bad, if the number of observation is low. That means when the observation is small, the Henderson's method 3 and its modified are limited to apply.

One of the variance components σ_1^2 must be seen as the main interest. The core procedure of modified Henderson's method is focused on the estimation for σ_1^2. And the constants of $\hat{\sigma}_{11}^2$ and $\hat{\sigma}_{12}^2$ used to modify are determined by minimizing the leading terms of $MSE\left(\hat{\sigma}_{u1}^2\right)$ and $MSE\left(\hat{\sigma}_1^2\right)$. In real experiments, we have to choose one of the random effects as the main interest if the modified Henderson's method 3 is applied to estimate the variance components. We can not expect to have the improving estimator for all the variance components. So, when we focus on all the variance components, it is not suitable to consider modified Henderson's method 3.

A APPENDICES

APPENDIX A

1. $Y = 1_8\mu + \begin{pmatrix} 1_4 & 0 \\ 0 & 1_4 \end{pmatrix} u_1 + \begin{pmatrix} 1_2 & 0 \\ 0 & 1_2 \\ 1_2 & 0 \\ 0 & 1_2 \end{pmatrix} u_2 + e$, $n = 8$, $p = 2$, and $q = 2$

2. $Y = 1_8\mu + \begin{pmatrix} 1_5 & 0 \\ 0 & 1_3 \end{pmatrix} u_1 + \begin{pmatrix} 1_2 & 0 \\ 0 & 1_3 \\ 1_1 & 0 \\ 0 & 1_2 \end{pmatrix} u_2 + e$, $n = 8$, $p = 2$, and $q = 2$

3. $Y = 1_8\mu + \begin{pmatrix} 1_6 & 0 \\ 0 & 1_2 \end{pmatrix} u_1 + \begin{pmatrix} 1_4 & 0 \\ 0 & 1_2 \\ 1_1 & 0 \\ 0 & 1_1 \end{pmatrix} u_2 + e$, $n = 8$, $p = 2$, and $q = 2$

4. $Y = 1_{21}\mu + \begin{pmatrix} 1_5 & 0 & 0 \\ 0 & 1_9 & 0 \\ 0 & 0 & 1_7 \end{pmatrix} u_1 + \begin{pmatrix} 1_2 & 0 & 0 \\ 0 & 1_3 & 0 \\ 0 & 1_1 & 0 \\ 0 & 0 & 1_8 \\ 1_4 & 0 & 0 \\ 0 & 1_3 & 0 \end{pmatrix} u_2 + e$, $n = 21$, $p = 3$,and $q = 3$

5. $Y = 1_{30}\mu + \begin{pmatrix} 1_{10} & 0 & 0 \\ 0 & 1_{15} & 0 \\ 0 & 0 & 1_5 \end{pmatrix} u_1 + \begin{pmatrix} 1_5 & 0 & 0 \\ 0 & 1_5 & 0 \\ 1_{10} & 0 & 0 \\ 0 & 1_5 & 0 \\ 0 & 1_2 & 0 \\ 0 & 0 & 1_3 \end{pmatrix} u_2 + e$ with $n = 30$, $p = 3$, and $q = 3$

6. $Y = 1_{30}\mu + \begin{pmatrix} 1_7 & 0 & 0 & 0 \\ 0 & 1_{12} & 0 & 0 \\ 0 & 0 & 1_6 & 0 \\ 0 & 0 & 0 & 1_5 \end{pmatrix} u_1 + \begin{pmatrix} 1_4 & 0 & 0 \\ 0 & 0 & 1_3 \\ 0 & 1_{10} & 0 \\ 0 & 0 & 1_2 \\ 1_2 & 0 & 0 \\ 0 & 1_4 & 0 \\ 1_5 & 0 & 0 \end{pmatrix} u_2 + e$, $n = 30$, $p = 4$, and $q = 3$

APPENDIX B

Three definitions about the bias and MSE are given by Wackerly, Mendenhall and Scheaffer (2002).

Definition 5 *let $\hat{\theta}$ be a point estimator for θ.Then $\hat{\theta}$ is an unbiased estimator if $E\left(\hat{\theta}\right) = \theta$. otherwise, $\hat{\theta}$ is said to be biased.*

Definition 6 *The bias of a point estimator $\hat{\theta}$ is given by $B\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$.*

Definition 7 *The mean square error of a estimator is $\hat{\theta}$ the expected value of $\left(\hat{\theta} - \theta\right)^2$:*

$$MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$$

If the $B\left(\hat{\theta}\right)$ denotes the bias, it can be shown that $MSE\left(\hat{\theta}\right) = D\left(\hat{\theta}\right) + B\left(\hat{\theta}\right)^2$ where $D\left(\hat{\theta}\right)$ denotes sample variance of $\hat{\theta}$.

Let us define $\hat{\sigma}^2$ as the estimator to estimate the true value σ^2. The expectation and variance of $\hat{\sigma}^2$ are denoted as $E\left(\hat{\sigma}^2\right)$ and $D\left(\hat{\sigma}^2\right)$ respectively. We denote a sample set of data as $\hat{\sigma}^2 = \left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots \hat{\sigma}_N^2\right)$ with $i = 1, 2, \ldots, N$.

Then the observed sample mean of $\hat{\sigma}^2$ is given:

$$mean\left(\hat{\sigma}^2\right) = \frac{1}{N} \sum \hat{\sigma}_i^2$$

which replaces with the $E\left(\hat{\sigma}^2\right)$.

The observed sample variance denoted as $S^2\left(\hat{\sigma}^2\right)$ is calculated as:

$$S^2\left(\hat{\sigma}^2\right) = \frac{1}{N-1} \sum \left(\hat{\sigma}_i^2 - mean\left(\hat{\sigma}^2\right)\right)^2$$

which replaces with $D\left(\hat{\sigma}^2\right)$.

Moreover, The estimated bias is

$$Bias\left(\hat{\sigma}^2\right) = mean\left(\hat{\sigma}^2\right) - \sigma^2.$$

According to the definition, the observed MSE of is:

$$MSE\left(\hat{\sigma}^2\right) = S^2\left(\hat{\sigma}^2\right) + Bias\left(\hat{\sigma}^2\right)^2$$

The observed negative probability used in our article is:

$$P\left(\hat{\sigma}^2 < 0\right) = \frac{P}{N}$$

where P is the numbers of negative estimates.

APPENDIX C

Design matrixes for fixed and random effects in section 4 are given.

X includes 6 columns, Z_1 and Z_2 have 16 and 4 columns respectively.

$$
X=\begin{pmatrix}
1&1&0&1&1&0\\
1&0&1&1&0&1\\
1&0&0&1&0&0\\
1&1&0&1&1&0\\
1&0&1&1&0&1\\
1&0&0&1&0&0\\
1&1&0&1&1&0\\
1&0&1&1&0&1\\
1&0&0&1&0&0\\
1&1&0&1&1&0\\
1&0&1&1&0&1\\
1&0&0&1&0&0\\
1&1&0&0&0&0\\
1&0&1&0&0&0\\
1&0&0&0&0&0\\
1&1&0&0&0&0\\
1&0&1&0&0&0\\
1&0&0&0&0&0\\
1&1&0&0&0&0\\
1&0&1&0&0&0\\
1&0&0&0&0&0\\
1&1&0&0&0&0\\
1&0&1&0&0&0\\
1&0&0&0&0&0
\end{pmatrix}
Z_1=\begin{pmatrix}
1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0\\
0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0\\
0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1\\
1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0\\
0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0\\
0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0\\
0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1
\end{pmatrix}
Z_2=\begin{pmatrix}
1&0&0&0\\
1&0&0&0\\
1&0&0&0\\
0&1&0&0\\
0&1&0&0\\
0&1&0&0\\
0&0&1&0\\
0&0&1&0\\
0&0&1&0\\
0&0&0&1\\
0&0&0&1\\
0&0&0&1\\
1&0&0&0\\
1&0&0&0\\
1&0&0&0\\
0&1&0&0\\
0&1&0&0\\
0&1&0&0\\
0&0&1&0\\
0&0&1&0\\
0&0&1&0\\
0&0&0&1\\
0&0&0&1\\
0&0&0&1
\end{pmatrix}
$$

APPENDIX D

This reduced model method provided by Kelly and Mathew (1994) is to derive estimators that are invariant with the changes of the means of Y based on model (1). The $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$ can also be obtained from this method.

In order to delete the effect of σ_2^2 and β, we define a $n \times t$ matrix K' where $t = n - b$, $b = rank\,(X)$. The matrix K' satisfies $K'K = I$ and $K'\,(X, Z_2) = 0$. The columns of K' are orthogonal vectors with each other and orthogonal to the columns of (X, Z_2). Let us define $u = K'Y$, then a new model is given:

$$u = U_1 u_1 + K'e \tag{D.1}$$

where $U_1 = K'Z_1$.

According to the assumptions given in the section 2.1,

$$E\,(u) = 0$$

$$D\,(u) = \sigma_1^2 V_1^* + \sigma_e^2 I_t$$

where $V_1^* = U_1'U_1$.

The ANOVA method is applied to estimate σ_1^2 and σ_e^2 with the new model (D.1). So the sums of squares SSR due to u_1 is

$$SSR_{u_1} = u'Pu \tag{D.2}$$

where $P = U_1\left(U_1'U_1\right)^{-}U$ and is also a a idempotent matrix.

The sum of squares due to e is

$$SSE = u^{'}\left(I - P\right)u \tag{D.3}$$

Based on the the properties of quadratic forms, we equate SSR_{u_1} and SSE to their expectation.

$$E\begin{pmatrix} SSR_{u_1} \\ SSE \end{pmatrix} = T\begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} \tag{D.4}$$

where $T = \begin{pmatrix} tr\left(PV_1^*\right) & tr\left(P\right) \\ 0 & tr\left(I - P\right) \end{pmatrix}$.
Then solving the equation (D.4), the solution of variance components is:

$$\begin{pmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_1^2 \end{pmatrix} = T^{-1}\begin{pmatrix} u^{'}Pu \\ u^{'}\left(I - P\right)u \end{pmatrix} \tag{D.5}$$

So the expression of $\hat{\sigma}_1^2$ is :

$$\hat{\sigma}_1^2 = \frac{u^{'}Pu}{tr\left(PV_1^*\right)} - \frac{tr\left(P\right)u^{'}\left(I - P\right)u}{tr\left(PV_1^*\right)tr\left(I - P\right)} \tag{D.6}$$

Therefore, $\hat{\sigma}_1^2$ and $\hat{\sigma}_{12}^2$ can also be obtained from the reduced model method.

R code for estimation

```
### Henderson method's and its modified
    HendersonEst<-function(Y,X,Z1,Z2) {
    ###############################################
    # This function calculates estimates of Henderson 3 and its modifiedof partition I and Partition II respectively.
    # The model used is mixed linear model with two random effects and iid residuals.
    # Code written by Lars Ronnegard 2008-07-29, modified by Weigang Qie.
    # Input:
    # Y = response vector
    # X = design matrix for fixed effects
    # Z1 = incidence matrix for first random effect
    # Z2 = incidence matrix for second random effect
    ###############################################
    library(MASS)
    n<-length(Y)
    ###############################################
    X1<-cbind(X,Z1)
    X2<-cbind(X1,Z2)
    PX<-X%*%ginv(t(X)%*%X)%*%t(X)
    PX1<-X1%*%ginv(t(X1)%*%X1)%*%t(X1)
    PX2<-X2%*%ginv(t(X2)%*%X2)%*%t(X2)
    I<-diag(1,length(Y))
    A<-PX1-PX
    B<-PX2-PX1
    C<-I-PX2
    V1<-Z1%*%t(Z1)
    V2<-Z2%*%t(Z2)
    a<-sum(diag(A%*%V1))
    b<-sum(diag(B%*%V2))
    c<-sum(diag(C))
    d<-sum(diag(A%*%V2))
    e<-sum(diag(B))
    #Added f to equations
    f<-sum(diag(A))
    #Used lower case k since it is a scalar
    k<-d*e-f*b
    ############################################
    ######### nonmodified Henderson III ##############
    ############################################
    ##Partition I
    sigma1<-round(1/a*(t(Y)%*%A%*%Y-d/b*(t(Y)%*%B%*%Y)+k/(b*c)*(t(Y)%*%C%*%Y)),4)
    ##Partition II
    X3<- cbind(X,Z2)
    PX3<- X3%*%ginv(t(X3)%*%X3)%*%t(X3)
    E<- PX2-PX3
    g<- sum(diag(E%*%V1))
    l<- sum(diag(E))
    sigma12<-round(t(Y)%*%E%*%Y/g-l*t(Y)%*%C%*%Y/(c*g),4)
    sigma1hat<-cbind(sigma1,sigma12)
    ###########################################
    ######### modified Henderson III #################
    ###########################################
    ##Partition I
```

```
c1<-1/(2/(a^2)*sum(diag(A%*%V1%*%A%*%V1))+1)
d1<-1/(2/(b^2)*sum(diag(B%*%V2%*%B%*%V2))+1)
d2<-(d/b*d1*sum(diag(B))-sum(diag(A)))/((k/b)*(2/c+1))
##############
modsigma1<-c1/a*(t(Y)%*%A%*%Y-d/b*d1*(t(Y)%*%B%*%Y)+k/(b*c)*d2*(t(Y)%*%C%*%Y))
##Partition II
c2<- g^2/(2*sum(diag(E%*%V1%*%E%*%V1))+g^2)
e1<-c/(2+c)
#############
modsigma12<-c2*t(Y)%*%E%*%Y/g-c2*e1*l*t(Y)%*%C%*%Y/(c*g)
#############
modsigma1hat<-cbind(modsigma1,modsigma12)
#############
r<-cbind(sigma1hat,modsigma1hat)
r
}
#####################################
######## Simulation function ###############
#####################################
simu2<-function(Z1,Z2,X,mu,n,a,b,sigma1,sigma2,sigmae){
###Calculate MSE of Partition I and PartitionII for nonmodified and modified Henderson
###Calculate MSE of ML and REML
Re<-M<-h<-h2<-w<-m<-numeric(100)
i<-1
while(i<=100){
b1<-as.matrix(c(1:a))
b2<-as.matrix(c(1:b))
c1<-as.factor(c(Z1%*%b1))
c2<-as.factor(c(Z2%*%b2))
e<-as.matrix(rnorm(n,0,sqrt(sigmae)))
u1<-as.matrix(rnorm(a,0,sqrt(sigma1)))
u2<-as.matrix(rnorm(b,0,sqrt(sigma2)))
Y<- X*mu+Z1%*%u1+Z2%*%u2+e
lm2<-lmer(Y~1+(1|c1)+(1|c2),REML=FALSE)
lm1<-lmer(Y~1+(1|c1)+(1|c2))
z<-VarCorr(summary(lm1))
x<-VarCorr(summary(lm2))
Re[i]<-as.numeric(as.matrix(z[1]))
M[i]<-as.numeric(as.matrix(x[1]))
s<-HendersonEst(Y,X,Z1,Z2)
h[i]<-s[1]
h2[i]<-s[2]
m[i]<-s[3]
w[i]<-s[4]
i<-i+1
}
M1<-round(var(h)+(mean(h)-sigma1)^2,4)
M2<-round(var(h2)+(mean(h2)-sigma1)^2,4)
Mo1<-round(var(m)+(mean(m)-sigma1)^2,4)
Mo2<-round(var(w)+(mean(w)-sigma1)^2,4)
REML<-round(var(Re)+(mean(Re)-sigma1)^2,4)
ML<-round(var(M)+(mean(M)-sigma1)^2,4)
r<-rbind(M1,M2,Mo1,Mo2,REML,ML)
r
```

```
}
###########################
####### Data application ########
##########################
## Data description
y<-c(56,50,39,30,36,33,32,31,15,30,35,17,41,36,35,25,28,30,24,27,19,25,30,18)
hist(y,freq=F,xlab='Y',ylab='denstiy',main='Distribuion of Y')
lines(density(y),lwd=2)
min(y)
max(y)
mean(y)
var(y)
r<-seq(15,56,len=24)
d<-dnorm(r,mean(y),sqrt(var(y)))
lines(r,d,lty=2,lwd=2)
legend(40,0.049,c(expression(Y),'normal distribution'),lty=1:2,lwd=c(2,1))
qqnorm(y,main='Normal Q-Q Plot of Y')
qqline(y)
## Estimation
######################
HendersonEst(Y,X,Z1,Z2)
#####################
library(lme4)
lm<-lmer(y~a+b+a*b+(1|t)+(1|block))
lm2<-lmer(y~a+b+a*b+(1|t)+(1|block),REML=FALSE)
REML<-as.numeric(as.matrix(VarCorr(summary(lm))[1]))
r<-VarCorr(summary(lm2))
ML<-as.numeric(as.matrix(VarCorr(summary(lm2))[1]))
#####################
```

References

[1] Ahrens, H. and Pincus, R. (1981), On Two Measures of Unbalalanceness in a One-Way Model and Their Relation to Efficiency, Biometrics, Vol.23, NO.3, pp. 227-235

[2] Al-Sarraj, R. and Rosen, D.v (2008) Improving Henderson's 3 approach when estimating variance components in a two-way mixed linear model, Centre of Biostochastics, Swedish University of Agricultural Scicence, report no. 2007:8

[3] Christensen, R., Pearson, L. And Johnson, W. (1992), Case Deletion Diagnostics for Mixed Models, Technometrics 38-45

[4] Corbeil R.R. and Searle .S:R (1976), A Comparison of Variance Components Estimators, Biometrics, Vol.32, No.4, pp.779-791

[5] Davidson, A.C. (2003), Statistical Models, Cambridge University Press

[6] Giampaoli, V. and Singer, J.M. (2009), Likelihood ratio tests for variance components in linear mixed models, Journal of Statistical Planning and Inference 139 (2009) 1435 – 1448

[7] Graybill, F. A. and Hultquist, R.A. (1961), Theorems Concerning Eisenhart's Model II. Ann. Math. Stat. 32, 261-269

[8] Kelly, R.J. and Mathew, T. (1994), Improved Nonnegative Estimation of Variance Components in Some Mixed Models with Unbalanced Data, Technometrics, Vol. 36, No. 2, pp. 171-181

[9] Khuri, A. I., Mathew, T., Sinha, B. K.(1998). Statistical Tests for Mixed Linear Modeles. John Wiley &sons, New York

[10] Harville D.A. (1977), Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems, Journal of American Statistical Association, Vol.72, No, 358, pp. 320-338

[11] Henderson, C.R. (1953), Estimation of variance and covariance components, Biometrics 9: 226-252.

[12] Herbach, L.H. (1959), Properties of Model II–Type Analysis of Variance Tests, A: Optimum Nature of the F-Test for Model II in the Balanced Case, The Annals of Mathematical Statistics, Vol. 30, No. 4, pp. 939-959

[13] Marsaglia, G. and Bray, T. A. (1964), A Convenient Method for Generating Normal Deviates, SIAM Review, 6, 260-264

[14] McCulloch, C.E. and Searle, S. (2002), Generalized Linear and Mixed Models, Wiley:NY.

[15] Patterson, H. D. and Thompson, R. (1971), Recovery of Inter-Block Information when Block Sizes are Unequal, Biometrika, Vol. 58, No. 3, pp. 545-554

[16] Rao, R. (1972), Estimation of Variance and Covariance components in linear models, Journal of American Statistical Association, Vol.67, No, 337, pp. 112-115

[17] Ramon L.C., (1996), SAS system for mixed models, Cary, NC : SAS Inst.

[18] Searle, S.R. (1987) Linear models for unbalanced data. John Wiley &sons, New York

[19] Searle, S.R., Casella, G and McCulloch, C.E. (1992), Variance Components, Wiley:NY. New York

[20] Swallow, W.H. and Searle, S.R (1978), Minimum Variance Quadratic Unbiased Estimation (MIVQUE) of Variance components, Technometrics 20,165-172

[21] Swallow, W.H. and Monahan, J.F. (1984), Monte Carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components, Technometrics, Vol. 26, No. 1, pp. 47-57

[22] Wackerly, D, D., Mendenhall, W. and Scheaffer, R.L. (2002), Mathematical Statistics with applications, Duxbury press 2002, ISBN 0-534-37741-6.