On Estimating Variance Components of Two-Way Nested Random Model with Missing Information

Amany Mousa Mohammed[1]; Ahmed Amin El-Sheikh[1]; Alaa Sayed Shehata[2*]

[1]Professor of Applied Statistics and Econometrics, Institute of Statistical Studies and Research, Cairo University

[2]Ph.D student of Applied Statistics and Econometrics, Institute of Statistical Studies and Research, Cairo University

*Alaa_sayed_sh@yahoo.com

ABSTRACT:

In this paper, the estimators of variance components are derived of two-way nested random model when the problem of missing information exists using combination between Modified Minimum Variance Quadratic Unbiased Estimation (MMIVQUE) and Modified Minimum Variance Quadratic Unbiased Estimation (MMIVQUE (0)) methods that is called MMIV(MIV(0)) method.

Keywords: MINQUE, Missing Information, MIVQUE, Variance Components.

INTRODUCTION

The problem of estimation of variance components in random and mixed linear models has received much attention in the statistics literature. There are several approaches to this problem, such as the analysis of variance (ANOVA) estimator. It has been common practice estimate the variance components by ANOVA for balanced data. The ANOVA estimates are obtained by equating observed and expected mean squares in the analysis and solving the resulting equation for the estimators. These estimators are unbiased and can be expressed as quadratic functions of the observations. The main desirable feature of these estimators is their simple computation. Under normality and balanced data, they have minimum variance among all unbiased estimators. However they can yield negative estimates and even under normality assumptions their distributions are intractable. For unbalanced data, the choice of appropriate quadratic forms poses a difficult problem. The estimates obtained may be not unbiased. (Li, 1995)

Rao (1970) suggested a method of estimation “MINQUE” that does not require the normality assumption for the estimation of variances. Rao (1971) proposed a method of estimation that called MIVQUE, Minimum Variance Quadratic Unbiased Estimation. Swallow and Monahan (1984) made a comparison between ANOVA, MLE, REML, and MINQUE methods through running one way model
Subramani (2012) suggested a modification on the computational aspects of MIVQUE of variance components in mixed linear models. He introduced two modified MIVQUE (MIVQUE I and MIVQUE II). He estimated variance components in unbalanced one-way random model by Modified MIVQUE and compared between MIVQUE I, MIVQUE II, MIVQUE based on different optimality criteria.

Most standard statistical methods have been designed to analyze data sets with no missing values. Consequently, the researcher has two options (a) to delete those cases which have missing data, or (b) to fill-in the missing values with estimated values. Thus, a data set is created containing no missing values (empty cells). Typically, the data set is presented in a rectangular table where rows indicate cases, observations, or subjects, and columns indicate variables measured on each unit.
In regression analysis, independent variables may have missing values in practice. It is also likely that information (which group or subgroup an observation belongs to) in the analysis of variance is missing.

The information in variance component model has the same importance as the independent variables in regression analyses. Without the information, the variance components in the model cannot be separated from one another (i.e. it may make some variance components inestimable). (Song and Shulman, 1997)

The meaning of incomplete (missing) information is different from the meaning of missing values. Missing values related to the losing of the observation while missing information related to the losing of location of the observation. This means that the value of the observation is known which could happen because it may be not recorded or lost for any other reasons. Missing information has three types completely missing information, partially missing information and not at all on any observation.

1. Completely missing information
   No observation in the main group has subgroup information.
2. Partially missing information
   Some of the observations in the main group have missing subgroup information.
3. Not at all on any observation in the main group is missing. (Saleh and El-sheikh, 2002)

Song and Shulman (1997) estimated the variance components for the data with missing nesting information in the two-stage unbalanced nested random model. They combined sum of squares for the data with missing nesting information with the sum of squares for the data with complete nesting information linearly. Prespecified weights are used for the combination. Different estimates are obtained by using different weights. Variances and covariances of these estimators are derived and used to compare these estimators. Saleh and El Sheikh (2003) modified the analysis of variance method and the combined symmetric sums with the analysis of variance method for estimation of the variance components of three-stage unbalanced nested random models for the data with complete missing nesting information. By a simulation study, they compared the bias and the mean squares errors of the estimates of variance components of the five methods of estimation namely: ANOVA method (Henderson's method 1), Modified ANOVA method, Combined analysis of means with ANOVA method, Combined symmetric sums method with ANOVA method, Combined symmetric sums method with modified ANOVA method.
The paper is organized as follows: The second section concerns with the Modified MIVQUE (I) method introduced by Subramani (2012). The third section illustrates the proposed estimators for data with completely missing information in case of two-way nested random model. The fourth section illustrates simulation study to compare ANOVA and MMIV(MIV(0)) methods.

MINIMUM VARIANCE QUADRATIC UNBIASED ESTIMATION (MIVQUE I)

Assume the model:

$$Y = X\beta + Z_1\delta_1 + Z_2\delta_2 + \cdots + Z_d\delta_d \quad (1)$$

where Y is an $N \times 1$ vector of observations, N is the sample size
X is a $N \times s$ matrix with known constants,
β is a $s \times 1$- vector of fixed (unknown) parameters,
Z_i is a $N \times c_i$ matrix with known constants, $i = 1, \dots, d$. $(Z_d = I, c_d = N)$
δ_i is a $c_i \times 1$-vector of random variables. $(\delta_d = e)$
Assume that δ_i is random variable with zero mean value and dispersion matrix $\sigma_i^2 I_{c_i}$. Further, δ_i and δ_j are uncorrelated.
Model (1) can be expressed in a compact form as:

$$Y = X\beta + Z\delta \quad (2)$$

where $Z = (Z_1 : \ Z_2 : \cdots : Z_d)$ and $\delta = (\delta_1 : \ \delta_2 : \cdots : \delta_d)$.
$E(Y) = X\beta$ and $D(Y) = V = \sum_{i=1}^{d} \sigma_i^2 V_i$ where $V_i = Z_i Z_i^t$.
$D(Y)$ is called the dispersion matrix and the parameters $\sigma_1^2, \dots, \sigma_d^2$ are the unknown variance components whose values should be estimated. (Subramani, 2012)
Subramani (2012) developed the estimation of variance components based on Rao (1971) approach. Instead of dealing with one linear combination, he decided to estimate a set of linear combinations of variance components $\sum_{j=1}^{d} \rho_{ij}\sigma_i$ through a set of quadratic functions $Y^t A_i Y$ $[A_i \text{ is a symmetric matrix and } \rho_{ij} = Tr(A_i V_j)]$.

He claimed that estimating variance components under normality obtained by solving the following equations:

$$\begin{bmatrix} Tr(A_1V_1) & \cdots & Tr(A_1V_d) \\ \vdots & \ddots & \vdots \\ Tr(A_dV_1) & \cdots & Tr(A_dV_d) \end{bmatrix}_{d\times d} \begin{bmatrix} \sigma_1^2 \\ \vdots \\ \sigma_d^2 \end{bmatrix}_{d\times 1} = \begin{bmatrix} Tr(A_1W) \\ \vdots \\ Tr(A_dW) \end{bmatrix}_{d\times 1} \quad (3)$$

He introduced different formulas of A_i to obtain MIVQUE(I). The formulas of A_i $(i = 1,2,\dots,d)$ have the following form:

$A_i = V^{-1}(I - P_{U_i})$ where $P_{U_i} = U_i(U_i^tV^{-1}U_i)^-U_i^tV^{-1}$. U_i has a variety of choices,

i. $U_1 = X, U_2 = [X \quad Z_1], U_3 = [X \quad Z_2], \dots, \; U_d = [X \quad Z_{d-1}]$

ii. $U_1 = X, U_2 = [X \quad Z_1], U_3 = [X \quad Z_1 \quad Z_2], \dots,$
$U_d = [X \quad Z_1 \quad Z_2 \quad \dots \quad Z_{d-1}]$ (4)

iii. $U_1 = X, U_2 = [X \quad X \quad Z_1], U_3 = [X \quad X \quad Z_1 \quad Z_2], \dots,$
$U_d = [X \quad X \quad Z_1 \quad \dots \quad Z_{d-1}]$

where $(U_i^tV^{-1}U_i)^-$ is the generalized inverse of $U_i^tV^{-1}U_i$

For the case (ii), he derived the estimators, their variances and covariance matrix in the unbalanced one-way random model. The resulting method are referred to as MIVQUE I.

The proposed estimators of variance components are derived by replacing A_i in eq. (3) by A_i for case (iii) in eq. (4).

So the steps of MIVQUE method: 1- Selecting a symmetric matrix A_i, 2- Solving the equation (3), 3-obtain the estimators of MIVQUE method.

ESTIMATION OF VARIANCE COMPONENTS FOR DATA WITH COMPLETELY MISSING INFORMATION

In this section, the variance components will be estimated for data with completely missing information by combination between modified MIVQUE I and modified MIVQUE I(0).

Consider the two- way nested random model

$$Y_{ijk} = \eta + \gamma_i + \beta_{j(i)} + e_{k(ij)} \quad (5)$$
$$i = 1,2,\dots,S, j = 1,2,\dots,D_i, k = 1,2,\dots,n_{ij}$$

where Y_{ijk} is the k^{th} observation at the j^{th}level of factor β within the i^{th} level of factor γ.

η is the general mean.

γ_i , $\beta_{ij}$and e_{ijk} are mutually independent random variables with zero means and variances σ_γ^2 , $\sigma_\beta^2$and σ_e^2 respectively. The variance components to be estimated are σ_γ^2 , σ_β^2 and σ_e^2.

So the model (5) can be written in matrix form as:

$$Y = X\eta + T_1\gamma + T_2\beta + T_3e \quad (6)$$

where Y is an $N \times 1$ vector of observations $X = 1_{N*1}, N = \sum_{i=1}^{S}\sum_{j=1}^{D_i} n_{ij}$

$$T_1 = \begin{bmatrix} 1_{\sum_{j=1}^{D_1} n_{1j}\times 1} & 0 & 0 & 0 & 0 \\ 0 & 1_{\sum_{j=1}^{D_2} n_{2j}\times 1} & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1_{\sum_{j=1}^{D_S} n_{Sj}\times 1} \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 1_{n_{11}\times 1} & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 1_{n_{1D_1}\times 1} & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1_{n_{SD_S}\times 1} \end{bmatrix}$$

$$, T_3 = I_N,$$

with $E(Y) = X\eta$ and $D(Y) = V = V_1\sigma_\gamma^2 + V_2\sigma_\beta^2 + V_3\sigma_e^2$

$$, V_i = T_i T_i^t.$$

$$V_1 = T_1 T_1^t = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & J_S \end{bmatrix} = \sum_{i=1}^{+S} J_i$$

$$V_2 = T_2 T_2^t = \begin{bmatrix} K_{n_{11}} & 0 & \dots & 0 \\ 0 & K_{n_{12}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_{n_{SD_S}} \end{bmatrix} = K$$

where J_i denote $\sum_{j=1}^{D_i} n_{ij} \times \sum_{j=1}^{D_i} n_{ij}$ matrix consisting of $1's$.

$K_{n_{ij}}$ denote $n_{ij} \times n_{ij}$ matrix consisting of $1's$.

Assume that the total number of the main group will be: $S = S' + S''$.

S' be the numbers of the main groups for the data with complete nesting information and S'' be the numbers of the main groups for the data with completely missing subgroup nesting information. Assume that all $D_i's$ and $n_{ij}'s$ in Model (4.1) are known.

Variables and coefficients without prime-notation or with single or double prime notations will be defined as follows:

If there is a notation without prime then we do not specify the range for i if the variable or coefficient is summed over i.

The same notation with a single prime (double primes) is then defined as the same quantity summed over i from, 1 to S' (from, $S'+1, \dots, S$, respectively).

Steps of estimation:

1. Estimation of variance components for S' groups. (data with complete information)
2. Estimation of variance components for S'' groups. (data with missing information)
3. Pre-specified weights will be used to combine data with complete information and missing information.

According to steps of MMIVQUE method, the estimators of variance components are derived when the matrix A_i for case (iii) in eq.(4).

For model (6), the matrix A_i is defined as:

$$A_i = V^{-1}\left(I - P_{U_i}\right), \qquad i = 1,2,3$$

where $P_{U_i} = U_i(U_i V^{-1} U_i)^{-} U_i^t V^{-1}$,

$U_1 = X, U_2 = [X \quad X \quad T_1], \widetilde{U}_{3U} = [X \quad X \quad T_1 \quad T_2]$

$$V^{-1}=\frac{1}{\sigma_e^2}I-\sum_{i=1}^{+S}B_i-\sum_{i=1}^{+S}\frac{\sigma_\gamma^2}{1+\sigma_\gamma^2\sum_{j=1}^{D_i}\frac{n_{ij}}{(\sigma_e^2+n_{ij}\sigma_\beta^2)}}C_i$$

where

$$B_i=\begin{bmatrix}\frac{\sigma_\beta^2}{\sigma_e^2(\sigma_e^2+n_{i1}\sigma_\beta^2)}K_{n_{i1}\times n_{i1}} & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \cdots & \vdots\\ 0 & 0 & \cdots & \frac{\sigma_\beta^2}{\sigma_e^2(\sigma_e^2+n_{iD_i}\sigma_\beta^2)}K_{n_{iD_i}\times n_{iD_i}}\end{bmatrix}$$

and

$$C_i=\begin{bmatrix}\frac{1}{(\sigma_e^2+n_{i1}\sigma_\beta^2)}\\ \vdots\\ \frac{1}{(\sigma_e^2+n_{iD_i}\sigma_\beta^2)}\end{bmatrix}\begin{bmatrix}\frac{1}{(\sigma_e^2+n_{i1}\sigma_\beta^2)} & \cdots & \frac{1}{(\sigma_e^2+n_{iD_i}\sigma_\beta^2)}\end{bmatrix}$$

By using MMIVQUE method, the variance components σ_γ^2 , $\sigma_\beta^2$and σ_e^2 will be replaced with the prior values α_1 , α_2 and α_0 respectively. So the dispersion matrix will take the following form:

$$V^*=\alpha_1V_1+\alpha_2V_2+\alpha_0V_3$$

So the inverse V^{-1} will be replaced with:

$$V^{*(-1)}=\frac{1}{\alpha_0}I-\sum_{i=1}^{+S}B_i-\sum_{i=1}^{+S}\frac{\alpha_1}{1+\alpha_1\sum_{j=1}^{D_i}\frac{n_{ij}}{(\alpha_0+n_{ij}\alpha_2)}}C_i$$

<u>Step (1): data with missing information by MMIVQUE I:</u>

For model (6), the matrix A_i', $i=1,2,3$ for data with complete information is:

$$A_1'=V_{(c)}^{*(-1)}-h'\left[V_{(c)}^{*(-1)}X'X'^tV_{(c)}^{*(-1)}\right]$$

$$A_2'=V_{(c)}^{*(-1)}-\left[\sum_{i=1}^{+S'}\frac{1}{h_i\left(1+\alpha_1\sum_{j=1}^{D_i}\frac{n_{ij}}{\alpha_0+n_{ij}\alpha_2}\right)^2}C_i\right]$$

$$A_{3U}'=V_{(c)}^{*(-1)}-\left[\sum_{i=1}^{+S'}F_i-\sum_{i=1}^{+S'}\frac{\alpha_1}{\left(1+\alpha_1\sum_{j=1}^{D_i}\frac{n_{ij}}{\alpha_0+n_{ij}\alpha_2}\right)}C_i\right]$$

where

$$F_i=\begin{bmatrix}\frac{1}{n_{i1}(\alpha_0+n_{i1}\alpha_2)}K_{n_{i1}\times n_{i1}} & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \cdots & \vdots\\ 0 & 0 & \cdots & \frac{1}{n_{iD_i}(\alpha_0+n_{iD_i}\alpha_2)}K_{n_{iD_i}\times n_{iD_i}}\end{bmatrix}$$

$$h'=\frac{1}{\sum_{i=1}^{S'}h_i}$$

The resulting equations are:

$$\begin{bmatrix}Tr(A_1'V_1') & Tr(A_1'V_2') & Tr(A_1'V_3')\\ Tr(A_2'V_1') & Tr(A_2'V_2') & Tr(A_2'V_3')\\ Tr(A_3'V_1') & Tr(A_3'V_2') & Tr(A_3'V_3')\end{bmatrix}\begin{bmatrix}\sigma_\gamma^2\\ \sigma_\beta^2\\ \sigma_e^2\end{bmatrix}=\begin{bmatrix}Q_1'\\ Q_2'\\ Q_3'\end{bmatrix}$$

where $Q_i'=Tr(A_i'W), W'=Y'Y'^t, i=1,2,3$

$$Q'_1 = \left[\frac{1}{\alpha_0}\sum_{i=1}^{S'}\sum_{j=1}^{D_i}\sum_{k=1}^{n} Y_{ijk}^2 - \sum_{i=1}^{S}\sum_{j=1}^{D_i}\frac{\alpha_2}{\alpha_0 a_{ij}}Y_{ij.}^2 - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{\alpha_1}{a_{ij}^2 b_i}Y_{ij.}^2 - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\sum_{\substack{l=1\\ j\neq l}}^{D_i}\frac{\alpha_1}{a_{ij}b_i a_{il}}Y_{ij.}Y_{il.}\right]$$

$$- h'\left[\frac{1}{\alpha_0}\sum_{i=1}^{S'}\sum_{j=1}^{D_i}\sum_{k=1}^{n} Y_{ijk} - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{n_{ij}\alpha_2}{\alpha_0 a_{ij}}Y_{ij.} - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\sum_{\substack{l=1\\ j\neq l}}^{D_i}\frac{\alpha_1 n_{ij}}{b_i}\left(\frac{1}{a_{ij}^2}Y_{ij.} + \frac{1}{a_{ij}a_{il}}Y_{il.}\right)\right]^2$$

$$Q'_2 = \left[\frac{1}{\alpha_0}\sum_{i=1}^{S'}\sum_{j=1}^{D_i}\sum_{k=1}^{n} Y_{ijk}^2 - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{\alpha_2}{\alpha_0 a_{ij}}Y_{ij.}^2 - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{\alpha_1}{a_{ij}^2 b_i}Y_{ij.}^2 - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\sum_{\substack{l=1\\ j\neq l}}^{D_i}\frac{\alpha_1}{a_{ij}b_i a_{il}}Y_{ij.}Y_{il.}\right]$$

$$- \left[\sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{1}{h_i a_{ij}^2 b_i^2}Y_{ij.}^2 + \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\sum_{\substack{l=1\\ j\neq l}}^{D_i}\frac{1}{h_i a_{ij} b_i^2 a_{il}}Y_{ij.}Y_{il.}\right]$$

$$Q'_3 = \frac{1}{\alpha_0}\left[\sum_{i=1}^{S'}\sum_{j=1}^{D_i}\sum_{k=1}^{n} Y_{ijk}^2 - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{1}{n_{ij}}Y_{ij.}^2\right]$$

and

$$\mathrm{Tr}(A'_1V'_1) = \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{n_{ij}}{a_{ij}b_i} - h'\sum_{i=1}^{S'}(h_i)^2$$

$$\mathrm{Tr}(A'_1\widetilde{V}'_2) = \left[\sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{n_{ij}[a_{ij}b_i - n_{ij}\alpha_1]}{a_{ij}^2 b_i}\right] - h'\sum_{i=1}^{S'}\sum_{i=1}^{D_i}\sum_{\substack{l=1\\ j\neq l}}^{D_i} n_{ij}\left[\frac{n_{ij}(b_i a_{ij} - n_{ij}\alpha_1)}{a_{ij}^3 b_i^2} - \frac{n_{il}^2\alpha_1}{a_{ij}b_i^2 a_{il}^2}\right]$$

$$\mathrm{Tr}(A'_1V'_3) = \sum_{i=1}^{S'}\sum_{j=1}^{D_i} n_{ij}\left[\frac{1}{\alpha_0} - \frac{\alpha_2}{\alpha_0 a_{ij}} - \frac{\alpha_1}{b_i a_{ij}^2}\right] - h'\sum_{i=1}^{S'}\sum_{i=1}^{D_i}\sum_{\substack{l=1\\ j\neq l}}^{D_i} n_{ij}\left[\frac{(b_i a_{ij} - n_{ij}\alpha_1)}{a_{ij}^3 b_i^2} - \frac{n_{il}\alpha_1}{a_{ij}b_i^2 a_{il}^2}\right]$$

$$\mathrm{Tr}(A'_2V'_2) = \left[\sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{n_{ij}[a_{ij}b_i - n_{ij}\alpha_1]}{a_{ij}^2 b_i}\right] - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{n_{ij}^2}{a_{ij}^2 b_i^2 h_i}$$

$$\mathrm{Tr}(A'_2V'_3) = \sum_{i=1}^{S'}\sum_{j=1}^{D_i} n_{ij}\left[\frac{1}{\alpha_0} - \frac{\alpha_2}{\alpha_0 a_{ij}} - \frac{\alpha_1}{b_i a_{ij}^2}\right] - \sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{n_{ij}}{a_{ij}^2 b_i^2 h_i}$$

$$\mathrm{Tr}(A'_3V'_3) = \sum_{i=1}^{S'}\sum_{j=1}^{D_i} n_{ij}\left[\frac{1}{\alpha_0} - \frac{\alpha_2}{\alpha_0 a_{ij}} - \frac{\alpha_1}{b_i a_{ij}^2}\right] - \left[\sum_{i=1}^{S'}\sum_{j=1}^{D_i}\frac{[a_{ij}b_i - n_{ij}\alpha_1]}{a_{ij}^2 b_i}\right]$$

Since

$$\mathrm{Tr}(A'_1V'_1)\sigma_\gamma^2 + \mathrm{Tr}(A'_1V'_2)\sigma_\beta^2 + \mathrm{Tr}(A'_1V'_3)\sigma_e^2 = Q'_1 \quad (7)$$

$$\mathrm{Tr}(A'_2V'_2)\sigma_\beta^2 + \mathrm{Tr}(A'_2V'_3)\sigma_e^2 = Q'_2 \quad (8)$$

$$\mathrm{Tr}(A'_3V'_3)\sigma_e^2 = Q'_3 \quad (9)$$

By solving eq.(7), (8) and (9), the estimators of variance components for data with complete information can be given as follows:

$$\widetilde{\sigma}_e^2 = \frac{Q'_3}{\mathrm{Tr}(A'_3V'_3)}$$

$$\widetilde{\sigma}_\beta^2 = \frac{Q'_2 - \mathrm{Tr}(A'_2V'_3)\widetilde{\sigma}_e^2}{\mathrm{Tr}(A'_2V'_2)}$$

$$\widetilde{\sigma}_\gamma^2 = \frac{Q'_1 - \mathrm{Tr}(A'_1V'_3)\widetilde{\sigma}_e^2 - \mathrm{Tr}(A'_1V'_2)\widetilde{\sigma}_\beta^2}{\mathrm{Tr}(A'_1V'_1)}$$

Step (2): data with missing information by MMIVQUE I(0):

The estimators of variance components are derived by MMIVQUE I(0) for data with completely missing information i.e it is assumed that the initial values $\alpha_1 = \alpha_2 = 0$ and $\alpha_0 = 1$.

For model (6), the matrix $A''_{iU}, i = 1,2,3$ for data with completely missing information is:

$$A''_1 = V^{*(-1)}_{(m)} - h''\left[V^{*(-1)}_{(m)} X'' X''^{t} V^{*(-1)}_{(m)}\right]$$

$$A''_2 = V^{*(-1)}_{(m)} - \left[\sum_{i=S'+1}^{+S} \frac{1}{h_i\left(1 + \alpha_1 \sum_{j=1}^{D_i} \frac{n_{ij}}{\alpha_0 + n_{ij}\alpha_2}\right)^2} C_i\right]$$

$$A''_3 = V^{*(-1)}_{(m)} - \left[\sum_{i=S'+1}^{+S} F_i - \sum_{i=S'+1}^{+S} \frac{\alpha_1}{\left(1 + \alpha_1 \sum_{j=1}^{D_i} \frac{n_{ij}}{\alpha_0 + n_{ij}\alpha_2}\right)} C_i\right]$$

where

$$V^{*(-1)}_{(m)} = \frac{1}{\alpha_0} I_{N''} - \sum_{i=S'+1}^{+S} B_i - \sum_{i=S'+1}^{+S} \frac{\alpha_1}{1 + \alpha_1 \sum_{j=1}^{D_i} \frac{n_{ij}}{(\alpha_0 + n_{ij}\alpha_2)}} C_i$$

$$h'' = \frac{1}{\sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} n_{ij}}$$

The resulting equations are:

$$\begin{bmatrix} Tr(A''_1 V''_1) & Tr(A''_1 V''_2) & Tr(A''_1 V''_3) \\ Tr(A''_2 V''_1) & Tr(A''_2 V''_2) & Tr(A''_2 V''_3) \\ Tr(A''_3 V''_1) & Tr(A''_3 V''_2) & Tr(A''_3 V''_3) \end{bmatrix} \begin{bmatrix} \sigma^2_\gamma \\ \sigma^2_\beta \\ \sigma^2_e \end{bmatrix} = \begin{bmatrix} Q''_1 \\ Q''_2 \\ Q''_3 \end{bmatrix}$$

where

$$Q''_1 = \left[\sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} \sum_{k=1}^{n_{ij}} Y^2_{ijk}\right] - h''\left[\sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} \sum_{k=1}^{n_{ij}} Y_{ijk}\right]^2$$

$$Q''_2 = \left[\sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} \sum_{k=1}^{n_{ij}} Y^2_{ijk}\right] - \left[\sum_{i=S'+1}^{S} \frac{1}{n_{i.}} Y^2_{i..}\right]$$

$$Q''_3 = \left[\sum_{i=1}^{S''} \sum_{j=1}^{D_i} \sum_{k=1}^{n_{ij}} Y^2_{ijk} - \sum_{i=1}^{S''} \sum_{j=1}^{D} \frac{1}{n_{ij}} Y^2_{ij.}\right]$$

and

$$Tr(A''_1 V''_1) = \sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} n_{ij} - h'' \sum_{i=S'+1}^{S} \left(\sum_{j=1}^{D_i} n_{ij}\right)^2$$

$$Tr(A''_1 V''_2) = \left[\sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} n_{ij}\right] - h'' \sum_{i=S'+1}^{S} \sum_{i=1}^{D_i} n^2_{ij}$$

$$Tr(A''_1 V''_3) = \sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} n_{ij} - h'' \sum_{i=S'+1}^{S} \sum_{i=1}^{D_i} n_{ij}$$

$$Tr(A''_2 V''_2) = \left[\sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} n_{ij}\right] - \sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} \frac{n^2_{ij}}{\sum_{j=1}^{D_i} n_{ij}}$$

$$Tr(A''_2 V''_3) = \sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} n_{ij} - \sum_{i=S'+1}^{S} \sum_{j=1}^{D_i} \frac{n_{ij}}{\sum_{j=1}^{D_i} n_{ij}}$$

$$\mathrm{Tr}(A_3''V_3'') = \sum_{i=S'+1}^{S}\sum_{j=1}^{D_i} n_{ij} - \left[\sum_{i=S'+1}^{S} D_i\right]$$

Since

$$\mathrm{Tr}(A_1''V_1'')\sigma_\gamma^2 + \mathrm{Tr}(A_1''V_2'')\sigma_\beta^2 + \mathrm{Tr}(A_1''V_3'')\sigma_e^2 = Q_1'' \quad (10)$$

$$\mathrm{Tr}(A_2''V_2'')\sigma_\beta^2 + \mathrm{Tr}(A_2''V_3'')\sigma_e^2 = Q_2'' \quad (11)$$

$$\mathrm{Tr}(A_3''V_3'')\sigma_e^2 = Q_3'' \quad (12)$$

Step (3): Combination the data with complete information and completely missing information:

Pre-specified weights will be used to combine (Q_1'&Q_1'') and (Q_2'&Q_2'') as:

$$\mathrm{Tr}(A_3'V_3')\hat{\sigma}_e^2 = Q_3' \quad (13)$$

$$w_2Q_2' + (1-w_2)Q_2'' = w_2\left[\mathrm{Tr}(A_2'V_2')\hat{\sigma}_\beta^2 + \mathrm{Tr}(A_2'V_3')\hat{\sigma}_e^2\right] + (1-w_2)\left[\mathrm{Tr}(A_2''V_2'')\hat{\sigma}_\beta^2 + \mathrm{Tr}(A_2''V_3'')\hat{\sigma}_e^2\right] (14)$$

$$\begin{aligned} w_1Q_1' + (1-w_1)Q_1'' &= w_1\left[\mathrm{Tr}(A_1'V_1')\hat{\sigma}_\gamma^2 + \mathrm{Tr}(A_1'V_2')\hat{\sigma}_\beta^2 + \mathrm{Tr}(A_1'V_3')\hat{\sigma}_e^2\right] \\ &+ (1-w_1)\left[\mathrm{Tr}(A_1''V_1'')\hat{\sigma}_\gamma^2 + \mathrm{Tr}(A_1''V_2'')\hat{\sigma}_\beta^2 + \mathrm{Tr}(A_1''V_3'')\hat{\sigma}_e^2\right] \quad (15) \end{aligned}$$

By solving eq.(13), (14) and (15), the estimators of variance components are:

$$\hat{\sigma}_e^2 = \frac{Q_3'}{\mathrm{Tr}(A_3'V_3')}$$

$$\hat{\sigma}_\beta^2 = \frac{w_2Q_2' + (1-w_2)Q_2'' - [w_2\mathrm{Tr}(A_2'V_3') + (1-w_2)\mathrm{Tr}(A_2''V_3'')]\hat{\sigma}_e^2}{[w_2\mathrm{Tr}(A_2'V_2') + (1-w_2)\mathrm{Tr}(A_2''V_2'')]}$$

$$\dot{\sigma}_\gamma^2 = \frac{w_1Q_1' + (1-w_1)Q_1'' - P_2\hat{\sigma}_e^2 - P_1\hat{\sigma}_\beta^2}{[w_1\mathrm{Tr}(A_1'V_1') + (1-w_1)\mathrm{Tr}(A_1''V_1'')]}$$

where

$$P_1 = [w_1\mathrm{Tr}(A_1'V_2') + (1-w_1)\mathrm{Tr}(A_1''V_2'')],$$
$$P_2 = [w_1\mathrm{Tr}(A_1'V_3') + (1-w_1)\mathrm{Tr}(A_1''V_3'')]$$

SIMULATION STUDY OF TWO –WAY NESTED RANDOM MODEL

In this section, the variance components are estimated for unbalanced two-way nested random model under normality assumption in case of data with completely missing information through a simulation study by MMIV(MIV(0)) and ANOVA methods and to compare the estimates using mean squared error, bias, and probability of getting negative estimates.

A numerical comparison for two-way nested random model requires a n-pattern, true values for the variance components $\sigma_\gamma^2, \sigma_\beta^2$ and σ_e^2, a priori values α_1, α_2 and α_0 for the variance components $\sigma_\gamma^2, \sigma_\beta^2$ and σ_e^2 respectively, percentage of completely missing information and the weights.

As stated by Song and Shulman (1997), the weights can be simply set to a constant or derived by some optimal procedures. They presented four procedures of weights:

1. Set $w_1 = w_2 = 1$. This gives the estimates using the data with complete nesting information only.
2. Set $w_1 = w_2 = \frac{1}{2}$. This method gives equal weight to the sums of squares associated with both the complete and the missing nesting information.
3. Select $w_1$and w_2, that minimize the variances of the combined sums of squares.
4. Select $w_1$and w_2, that minimize the variances of the derived estimates: $V(\hat{\sigma}_\beta^2)$ and $V(\hat{\sigma}_\alpha^2)$, respectively.

In this section, the procedures (2, 3, and 4) are considered through a simulation study.

By simulation study, 5000 independent random sample are generated. It is assumed that the sample size is 60 observations, number of main groups $S = 8$, number of subgroups and sample size of each subgroup as given in table (1), the true values $\sigma_e^2, \sigma_\beta^2, \sigma_\gamma^2$ as given in table (2). Percentage of missing information levels 25%, 50%, 75% and different weights (w_1, w_2) are considered.

Table (1): The Number of Subgroups for Unbalanced Two-Way Nested Random Model

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8
$n_{11}=3$	$n_{21}=3$	$n_{31}=3$	$n_{41}=5$	$n_{51}=3$	$n_{61}=2$	$n_{71}=3$	$n_{81}=3$
$n_{12}=5$	$n_{22}=2$	$n_{32}=4$	$n_{42}=4$	$n_{52}=2$	$n_{62}=4$	$n_{72}=2$	$n_{82}=2$
	$n_{23}=2$		$n_{43}=2$		$n_{63}=2$		$n_{83}=4$

Table (2): Variance Components Configurations Used in The Simulation of Two-Way Nested Random Model

σ_γ^2	σ_β^2	σ_e^2
0.1	0.1	1
1	0.1	1
2	0.1	1
0.1	1	1
1	1	1
2	1	1
0.1	2	1
1	2	1
2	2	1

Table (3): Comparison of ANOVA and MMIV(MIV(0)) estimates for unbalanced two-way nested random model-25% data with completely missing information based on compound MSE, compound absolute Bias and prob. of getting Negative estimates

σ_β^2	σ_γ^2	w_1	w_2	ANOVA			MMIV(MIV(0))		
				Compound MSE	Compound absolute Bias	Prob. Negative estimates	Compound MSE	Compound absolute Bias	Prob. Negative estimates
0.1	1	0.5	0.5	1.03	1.17	0.86	2.15	1.4	0.53
		0	0.48	0.96	1.11	0.8	1.23	1.23	0.67
		0	0.89	0.9	1.05	0.82	1.3	1.14	0.72
	2	0.5	0.5	4	2.24	0.85	2.97	1.76	0.54
		0	0.48	3.6	2.1	0.8	4.09	2.3	0.69
		0	0.89	3.55	2.04	0.82	3.7	1.98	0.72
1	0.1	0.5	0.5	0.74	0.91	0.86	4.5	2.35	0.52
		0.02	0.27	0.82	0.97	0.81	0.91	0.98	0.84
		0.1	0.57	0.8	0.97	0.82	1.16	1.19	0.6
	2	0.5	0.5	4.4	2.69	0.85	3.84	2.32	0.57
		0	0.27	4.34	2.64	0.81	5.54	2.97	0.87
		0	0.57	4.1	2.55	0.81	4.2	2.55	0.66
2	0.1	0.5	0.5	3.11	1.82	0.86	6.82	3.06	0.54
		0	0.24	3.26	1.9	0.8	3.17	1.8	0.87
		0.01	0.5	3.45	1.98	0.8	3.43	1.97	0.67
	1	0.5	0.5	3.84	2.55	0.86	5.42	2.78	0.57
		0	0.24	3.88	2.48	0.8	4.32	2.61	0.88
		0	0.5	3.96	2.54	0.8	3.96	2.49	0.68

According to simulation study for unbalanced two-way nested random model, a number of conclusions are drawn from the results for all tables of this model which are summarized in the following points:

<u>25% data with completely missing information:</u>

- In case of unbalanced two-way nested random model, it does not matter computing the estimates of σ_e^2 for MMIV(MIV(0)) and ANOVA methods because they are the same.
- It is reasonable to note that the compound MSE of ANOVA method is lower than MMIV(MIV(0)) method.
- The compound absolute bias of ANOVA method is lower than MMIV(MIV(0)) method.
- According to the probability of getting negative estimates, MMIV(MIV(0)) method is better than ANOVA method.

Table (4): Comparison of ANOVA and MMIV(MIV(0)) estimates for unbalanced two-way nested random model-50% data with completely missing information based on compound MSE, compound absolute Bias and prob. of getting Negative estimates

σ_β^2	σ_γ^2	w_1	w_2	ANOVA			MMIV(MIV(0))		
				Compound MSE	Compound absolute Bias	Prob. Negative estimates	Compound MSE	Compound absolute Bias	Prob. Negative estimates
0.1	1	0.5	0.5	1.11	1.23	0.86	1.06	1.13	0.61
		0.5	0.72	1.03	1.17	0.85	0.82	0.9	0.54
		0.56	0.92	1	1.14	0.85	0.7	0.88	0.57
	2	0.5	0.5	4.2	2.34	0.87	4.1	2.29	0.68
		0.52	0.72	4.1	2.27	0.85	3.46	2.09	0.57
		0.57	0.92	3.92	2.19	0.84	2.5	1.65	0.59
1	0.1	0.5	0.5	0.7	0.87	0.87	0.95	1.09	0.62
		0.41	0.48	0.7	0.86	0.86	0.8	0.99	0.65
		0.46	0.58	0.71	0.88	0.86	0.88	1.05	0.6
	2	0.5	0.5	4.5	2.72	0.86	4.46	2.67	0.72
		0.49	0.48	4.52	2.73	0.86	4.57	2.7	0.73
		0.54	0.58	4.47	2.71	0.86	4.24	2.6	0.69
2	0.1	0.5	0.5	2.93	1.74	0.87	3.03	1.86	0.68
		0.4	0.43	2.9	1.71	0.88	2.9	1.75	0.75
		0.45	0.5	2.91	1.73	0.86	2.94	1.8	0.7
	1	0.5	0.5	3.66	2.48	0.86	3.56	2.35	0.72
		0.44	0.43	3.59	2.44	0.88	3.7	2.4	0.78
		0.49	0.5	3.61	2.45	0.87	3.59	2.35	0.72

<u>50% data with completely missing information:</u>

- In case of unbalanced two-way nested random model, it does not matter computing the estimates of σ_e^2 for MMIV(MIV(0)) and ANOVA methods because they are the same.
- When $\sigma_\gamma^2 = 0.1$, the compound MSE of ANOVA method is lower than MMIV(MIV(0)) method. Also, The compound absolute bias of ANOVA method is lower than MMIV(MIV(0)) method.
- ANOVA and MMIV(MIV(0)) methods approach at high level of true values of variance components.
- According to the probability of getting negative estimates, MMIV(MIV(0)) method is better than ANOVA method.

Table (5): Comparison of ANOVA and MMIV(MIV(0)) estimates for unbalanced two-way nested random model-75% data with completely missing information based on compound MSE, compound absolute Bias and prob. of getting Negative estimates

σ_β^2	σ_γ^2	w_1	w_2	ANOVA			MMIV(MIV(0))		
				Compound MSE	Compound absolute Bias	Prob. Negative estimates	Compound MSE	Compound absolute Bias	Prob. Negative estimates
0.1	1	0.5	0.5	1.55	1.47	0.89	1.77	1.53	0.85
		0.98	0.9	1.11	1.21	0.79	18.44	3.31	0.55
		0.9	0.96	1.13	1.21	0.83	1.54	1.21	0.56
	2	0.5	0.5	2.16	2.69	0.89	5.66	2.81	0.88
		0.99	0.9	4.01	2.26	0.79	48.13	5.12	0.61
		0.95	0.96	4.09	2.27	0.82	6.67	2.36	0.56
1	0.1	0.5	0.5	0.7	0.85	0.9	0.78	0.89	0.86
		0.96	0.79	0.72	0.89	0.79	9.55	2.71	0.59
		0.84	0.69	0.67	0.85	0.82	1.16	1.16	0.67
	2	0.5	0.5	5.36	2.98	0.9	5.98	3.13	0.9
		0.99	0.79	4.38	2.66	0.79	64.24	6.02	0.67
		0.97	0.69	4.65	2.75	0.78	18.1	3.88	0.71
2	0.1	0.5	0.5	2.52	1.54	0.89	2.59	1.57	0.87
		0.98	0.77	2.83	1.73	0.78	32.48	4.82	0.64
		0.92	0.58	2.68	1.64	0.81	2.47	2.42	0.69
	1	0.5	0.5	3.56	2.41	0.9	3.83	2.48	0.89
		0.99	0.77	3.48	2.36	0.79	68.12	6.34	0.68
		0.97	0.58	3.54	2.36	0.79	21.45	4.19	0.69

75% data with completely missing information:

- In case of unbalanced two-way nested random model, it does not matter computing the estimates of σ_e^2 for MMIV(MIV(0)) and ANOVA methods because they are the same.
- The compound MSE of ANOVA method is lower than MMIV(MIV(0)) method. Also, the compound absolute bias of ANOVA methods are lower than MMIV(MIV(0)) method.
- According to the probability of getting negative estimates, MMIV(MIV(0)) method is better than ANOVA and methods.

CONCLUSION

The aim of this paper was to evaluate the performance of the proposed estimators relative to ANOVA's estimator via simulation studies. Different criteria such as mean square error, bias and probability of getting negative estimates are used to show the performance of the estimators under the study.

From simulation study, we estimated the variance components by MMIV(MIV(0)) and ANOVA methods under normality assumption and compared the estimators for unbalanced two-way nested random model. In unbalanced two-way random nested model, It is better to estimate variance component by MMIV(MIV(0)) method. ANOVA method has negative estimates which affects mean square error and bias.

REFERENCES:

1. Li, X. (1995), “Estimation of Variance Components with Missing Data”, Unpublished Ph.D thesis, Dalhousie University, Canada.
2. Rao C.R., Estimation of heteroscedastic variances in linear models, J. Amer. Stat. Assoc.,1970, 65, 112–115.
3. Rao C.R., Minimum Variance Quadratic Unbiased Estimation of Variance Component, Journal of Multivariate Analysis, 1971, 1, 445-456.
4. Sahai, H. and Ojeda, M.M., Analysis of Variance for Random Models, Unbalanced Data, Birkhauser, Boston, Basel, Berlin,2005.
5. Saleh, A.A. and El-Sheikh, A.A., Estimation of Variance Components for Three-Stage Unbalanced Nested Random Mixed Models with Incomplete Information, Proceeding of the 37th Annual Conference On Statistics, Computer Sciences and Operation Research,2002, 1-15, I.S.S.R, Cairo University, Egypt.
6. Saleh, A.A. and El-Sheikh, A.A., Estimation of Variance Components for Three-Stage Unbalanced Nested Random Mixed Models with Complete Missing Information (C.M.I), Proceeding of the 38th Annual Conference On Statistics, Computer Sciences and Operation Research,2003, 13-16, I.S.S.R, Cairo University, Egypt.
7. Song, R. and Shulman, S.A., Variance Components in the Two-Way Nested Model with Incomplete Nesting Information, Technometrics, 1997, 39(1), 71-80.
8. Subramani, J., On Modified Minimum Variance Quadratic Unbiased Estimation (MIVQUE) of Variance Components in Mixed Linear Models, model assisted statistics and applications, 2012,7, 179-200.
9. Swallow H, Monahan F. Monte-Carlo comparison of ANOVA, MINQUE, REML and ML estimators of variance components. Technometrics 1984, 26:47–57.