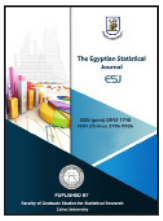
Cairo University

The Egyptian Statistical Journal

No.1 - Volume (67) , 2023

journal homepage: www.esju.journals.ekb.eg

Print ISSN 0542-1748 - Online ISSN 2786-0086



Impact of Negative Variance Component Estimates on the Kenward-Roger Test for Fixed Effects in Linear Mixed Models

Waseem Alnosaier

Department of Statistics, Institute of Public Administration,Riyadh,Saudi Arabia.

A B S T R A C T

A well-known procedure to make inference for fixed effects in a normal mixed linear model is the Kenward-Roger procedure (Kenward and Roger 1997), where a scaled Wald type statistic follows approximately an F distribution, and in special cases the test has an exact F test. In the procedure, the estimated denominator degrees of freedom of the F distribution, the estimated scale factor, and the scaled test statistic are calculated based on the data. The variance components of the random terms in the model are estimated by the restricted maximum likelihood (REML) method, in which sometimes the estimates produced are negative. Two methods are usually considered to resolve the issue of negative variance component estimates: to set the negative estimates as zero or to allow them remain negative. Assessing the performance of the procedure with each method based on the test level and power was done analytically and by a simulation study for four different designs. The estimates of the denominator degrees of freedom and scale factor were also calculated and compared for the procedure with each method. We showed that the Kenward-Roger procedure with the method that doesn't constrain the variance component estimates to be non-negative was instable for some data sets with negative variance component estimates, and was excessively conservative for some designs. By setting the negative variance component estimates as zero, the procedure became stable and more adequate than with allowing the estimates remain negative for the designs considered in the simulation study.

Keywords:

Variance components; Mixed linear model; Kenward Roger test; Restricted maximum likelihood estimator; Linear hypothesis.

1. **Introduction**

Inference for fixed effects in normal linear mixed models can be conducted using a test statistic that follows an exact F or t test when the data are balanced, and the random effects are not correlated. When the data is not balanced or the random effects are correlated, the test statistic follows an F distribution approximately. In both cases, the inference depends on the estimates of the variance components of the random effects. Most of variance components estimation procedures e.g., analysis of variance ANOVA, maximum likelihood (ML), and restricted maximum likelihood (REML), sometimes produce negative estimates (see Searl and Gruber 2016; Searl et al. 2006, chapter 2),

although variances should be positive by definition. The reasons of obtaining negative variance component estimates have been addressed by many researchers (see, e.g., Searl *et al.* 2006; Littell *et al.* 2006; Wu 1992; Zhang *et al.* 2000; Searl and Gruber 2016; Gao *et al.* 2008; Sjoberg 1995). Verdooren (1982) discussed the probability of having negative ANOVA estimates of the variance components and showed that this depends on the ratio of the variance components and the degrees of freedom in the ANOVA table. Since the variance components are positive by definition, a common practice among data analysts is to set the negative estimates as zero (Stroup and Littell 2002). However, setting the negative variance component estimates as zero requires attention and might affect the inference of fixed effects since it affects the estimates of other variance components, test statistic, and other estimates involved in the inference procedure. The impact of setting negative variance component estimates as zero and comparing this with allowing the estimates to remain negative for several procedures have been studies by several researchers. Stroup and Littell (2002) studied the impact of negative estimates of the variance components on the inference of fixed effects in unbalanced mixed linear models. They compared the performance of several procedures based on their rejection rates including the theoretical power proposed by Stroup (1999). However, studying the impact of negative variance component estimates on procedures that employs adjustment algorithms, e.g., Kenward-Roger (KR) procedure and the impact of negative estimates on the performance of the procedure has not received enough attention.

Based on the results of Kackar and Harville (1984), and Harville and Jeske (1992), Kenward and Roger (1997) improved the estimator of the variance of the fixed effects. Besides using the adjusted estimator of the variance of the fixed effects, Kenward and Roger developed their test by allowing the Wald-type test statistic to be scaled, and distributed approximately as F distribution under the null distribution. The scale factor of the test statistic and the denominator degrees of freedom of F distribution are estimated by matching moments, and the approximation is modified in such a way that the estimates match the known values for two special cases when the test is exact. Since, the KR procedure uses the REML method to estimate the variance components in the model, it is expected that the procedure sometimes produces negative variance components. In their simulation study, Kenward and Roger (1997) stated that the REML estimates of the variance components were not constrained to be non-negative. However, studying the effect of allowing the variance component estimates to remain negative on the procedure performance did not receive enough attention.

In this paper, we show that the KR procedure with negative variance component estimates as suggested by Kenward and Roger (1997) is not appropriate for all cases, and the procedure encounters a computational problem for some data sets where the procedure produces outlier estimates of the denominator degrees of freedom and the scale factor. On the other hand, we show that setting the estimates as zero for these data sets resolves the computational problem and makes the procedure perform adequately. In fact, when the only non-negative variance component estimate is the variance of the residual error, setting the negative variance estimates as zero makes the procedure not affected by the data characteristic, and the degrees of freedom and scale estimates produced are fixed. The method of setting the negative variance component estimates as zero will be called the “restricted” method, and the method of allowing the negative variance component estimates to remain negative will be called the “unrestricted” method. For the data sets with negative variance component estimates that the KR

procedure with the unrestricted method does not encounter any computational problems, we assess and compare the performance of the two methods of the KR procedure.

The model, notation, assumptions and the KR procedure are presented in section 2, and the issue of obtaining negative variance component estimates including some analytical results related to the KR procedure with the unrestricted and restricted methods are presented in section 3. Since the characteristic of the REML estimates of the variance components with both methods cannot be assessed analytically, a simulation study is presented in section 4 to assess and compare the performance of the KR procedure with the two methods. Four designs are considered in the simulation study: a balanced incomplete block design, a partial balanced incomplete block design, a random complete block design with missing data, and a crossover design, where the comparison is based on the level and power of the procedure with each method. The estimates of the denominator degrees of freedom, scale factors, and test statistics are calculated and compared for each method.

2. Testing the fixed effects in linear mixed models

Consider a multivariate normal data vector $\mathbf{y}$ with mean $\mathbf{X}\boldsymbol{\beta}$, and variance covariance $\boldsymbol{\Sigma}$ which is a linear function of unknown variance components $\boldsymbol{\sigma}$. $\mathbf{X}$ is a known $n \times p$ matrix of full column rank, and $\boldsymbol{\beta}$ is a $p \times 1$ matrix of the fixed effects parameters. Suppose that we are interested in testing the linear hypothesis $\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ where $\mathbf{L}$ is a fixed $p \times \ell$ matrix. Also, suppose that the test is performed by using the KR procedure as presented in Kenward and Roger (1997), and the test statistic is a Wald-type statistic of the form:

$$F = \frac{1}{\ell}\hat{\boldsymbol{\beta}}'\mathbf{L}[\mathbf{L}'\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\beta}})\mathbf{L}]^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}}, \tag{1}$$

where $\hat{\boldsymbol{\beta}}$ is the estimated generalized least-squares estimator (EGLSE) of $\boldsymbol{\beta}$, which is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \tag{2}$$

where $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})$, and $\hat{\boldsymbol{\sigma}}$ is the restricted likelihood estimator (REML) of the variance components $\boldsymbol{\sigma}$. Kenward and Roger improved the estimator of the $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ based on the results of Kackar and Harville (1984), and Harville and Jeske (1992) to be the adjusted estimator of $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ which is

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r}\hat{w}_{ij}(\hat{\mathbf{Q}}_{ij} - \hat{\mathbf{P}}_i\hat{\boldsymbol{\Phi}}\hat{\mathbf{P}}_j - \frac{1}{4}\hat{\mathbf{R}}_{ij})\right\}\hat{\boldsymbol{\Phi}}, \tag{3}$$

where

$\boldsymbol{\Phi} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$, $\mathbf{P}_i = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\boldsymbol{\Sigma}^{-1}\mathbf{X}$, $\mathbf{Q}_{ij} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\sigma_i}\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\sigma_j}\boldsymbol{\Sigma}^{-1}\mathbf{X}$, $\mathbf{R}_{ij} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\dfrac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\boldsymbol{\Sigma}^{-1}\mathbf{X}$, and these quantities to be estimated by substituting the REML estimates of the variance components, and $w_{ij} = \mathrm{Cov}(\hat{\sigma}_i, \hat{\sigma}_j)$ to be estimated by using the inverse of the observed information matrix. In addition, Kenward and Roger used a scaled form of F, say $F^* = \lambda F$ that is distributed approximately as $\mathrm{F}(\lambda, m)$ under the null hypothesis. Using Taylor series approximations and some convenient assumptions, and

matching the moments of F^* to the moments of an F distribution, they obtained formulas for both the denominator degrees of freedom *m*, and the scale factor λ.

$$m = 4 + \frac{\ell + 2}{\ell\varphi - 1}, \text{ and } \lambda = \frac{m}{E(m-2)} \tag{4}$$

where $\varphi = \frac{V}{2E^2}$.

E and *V* are the modified approximation of the expectation and variance of *F* so the approach produces the estimates of *m* and λ that match the exact known values for two special cases.

$$E = \left(1 - \frac{A_2}{\ell}\right)^{-1}, \text{ and } V = \frac{2}{\ell}\left\{\frac{1 + d_1 B}{(1 - d_2 B)^2 (1 - d_3 B)}\right\}, \tag{5}$$

where $B = \frac{1}{2\ell}(A_1 + 6A_2)$,

$A_1 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\,\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi})\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi})$, and $A_2 = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\,\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\boldsymbol{\Theta}\boldsymbol{\Phi}\mathbf{P}_j\boldsymbol{\Phi})$

for $\boldsymbol{\Theta} = \mathbf{L}(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}\mathbf{L}'$.

and

$d_1 = \frac{g}{3\ell + 2(1-g)}$

$d_2 = \frac{\ell - g}{3\ell + 2(1-g)}$

$d_3 = \frac{\ell - g + 2}{3\ell + 2(1-g)}$, for $g = \frac{(\ell+1)A_1 - (\ell+4)A_2}{(\ell+2)A_2}$.

3. The issue of negative estimates of variance components

The estimators of the variance components of the random effects in the model might be classified into two types: canonical and non-canonical estimators. While the canonical estimators are always non-negative, the non-canonical estimators e.g., Anova, maximum likelihood, and restricted maximum likelihood (REML) estimators, sometimes produce negative estimates. This happens because the calculation formulas used to produce the estimates do not ensure that the estimates are non-negative. Since the KR procedure uses the REML estimates of the variance components, sometimes the estimates of the variance components produced by the procedure are negative. A useful iteration form to compute the variance component estimates is equation (90) on p. 252 of Searl and *et. all*. (2006).

$$\left\{\mathrm{tr}\left(\mathbf{D}_i\mathbf{G}^{\bullet}\mathbf{D}_j\mathbf{G}^{\bullet}\right)\right\}_{i,j=1}^{r}\boldsymbol{\sigma}^{\bullet} = \left\{\mathbf{y}'\mathbf{G}^{\bullet}\mathbf{D}_i\mathbf{G}^{\bullet}\mathbf{y}\right\}_{i=1}^{r}, \tag{6}$$

where $\mathbf{G} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$,

and $\mathrm{Var}(\mathbf{y}) = \boldsymbol{\Sigma} = \sum_{i=1}^{r}\sigma_i^2\mathbf{D}_i$, for $\mathbf{D}_i$'*s* known symmetric matrices.

$\mathbf{G}^{\bullet}$ is $\mathbf{G}$ with the equation solution $\boldsymbol{\sigma}^{\bullet}$ used in place of $\boldsymbol{\sigma}$.

The iteration equation above to compute the REML estimates of the variance components clearly does not guarantee that the solutions $\boldsymbol{\sigma}^{\bullet}$ fall in the parameter space of the variance components $\boldsymbol{\sigma}$, and one or more of the variance component estimates might be negative based on the data characteristics. In fact, the problematic issue of obtaining negative variance component estimates can happen in theory for the estimates of σ_i^2 for $i \geq 2$, where the estimate of $\sigma_e^2 = \sigma_1^2$ cannot be negative. See the remark below.

Remark

The model presented in section 2 can be written in the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{7}$$

where $\mathbf{u}$ represents the random effects, $\mathbf{Z}$ is the corresponding matrix, and $\mathbf{e}$ is the residual errors.

Hartley and Rao (1967) developed two equations to calculate the maximum likelihood estimates of the variance components in terms of matrix $\mathbf{H}$, where $\boldsymbol{\Sigma} = \sigma_e^2\mathbf{H}$, and thus $\mathbf{H}$ is a function of the ratios σ_i^2 / σ_e^2 for all $i \geq 2$. Corbiel and Searle (1976) developed the equations for the REML estimators, and the iterative form to calculate the REML estimates of the residual errors is

$$\sigma_e^{2\bullet} = \frac{\mathbf{y}'\mathbf{K}'(\mathbf{K}\mathbf{H}^{\bullet}\mathbf{K}')^{-1}\mathbf{K}\mathbf{y}}{n-p}, \tag{8}$$

where $\mathbf{K}\mathbf{y}$ is the data vector used instead of the original data vector $\mathbf{y}$ to calculate the REML estimates of the fixed effects in the model, and $\mathbf{H}^{\bullet}$ is $\mathbf{H}$ with $\boldsymbol{\sigma}^{\bullet}$ used in place of $\boldsymbol{\sigma}$.

Notice that $\mathbf{K}'(\mathbf{K}\mathbf{H}\mathbf{K}')^{-1}\mathbf{K}$ is a non-negative definite matrix, and $\mathbf{y}'\mathbf{K}'(\mathbf{K}\mathbf{H}\mathbf{K}')^{-1}\mathbf{K}\mathbf{y}$ is a non-negative value (Harville 2008, theorem 14.2.9). Since typically the initial values of $\sigma_e^{2\bullet}$, and $\sigma_i^{2\bullet}$ for all $i \geq 2$ to perform the iteration are non-negative, then $\mathbf{y}'\mathbf{K}'(\mathbf{K}\mathbf{H}^{\bullet}\mathbf{K}')^{-1}\mathbf{K}\mathbf{y}$ will stay non-negative for all the iteration process, and hence the REML solution cannot be negative.

The analysis of the KR procedure might be performed by using the PROC MIXED procedure of SAS with adding option $\mathrm{DDFM} = \mathrm{KR}$ in the model statement (SAS Institute Inc 2021). Although the original KR procedure presented in Kenward and Roger (1997) did not constrain the REML estimates of the variance components to be non-negative, the PROC MIXED procedure set the negative estimates as zero by default. By using option "NOBOND" in the PROC MIXED procedure, SAS allows the negative variance component estimates to remain negative. However, as will be seen in the simulation study in section 4, the KR procedure with the unrestricted method encounters a computational problem for some data sets with negative variance component estimates based on the data characteristics. In particular, for these data sets with negative variance component estimates, the estimate of the quantity A_2 (as defined in section 2) is inflated in such a way that $A_2 > \ell$, and hence the obtained estimate of E (as defined in equation 5, section 2) becomes negative. This leads to outlier estimates of the denominator degrees of freedom, and the scale factor. For these data sets, the KR procedure with the unrestricted method, which is obtained in SAS by using option "NOBOND" in the PROC MIXED procedure, SAS only produces the variance component estimates, and does not perform any further

computations for the KR procedure. However, the KR procedure with the restricted method in which the negative variance component estimates set as zero, the PROC MIXED procedure performs the KR procedure and provides the result of the inference of the fixed effects including the estimates of the denominator degrees of freedom and the scale.

If the only non-negative variance component estimate is the estimate of the variance of the residual error, then the estimates of the denominator degrees of freedom and the scale factor produced by KR procedure with the restricted method are fixed and do not rely on the data characteristics. See the lemma below.

Lemma

For the model considered in section 2, and when all variance component estimates are negative except the variance of residual error, then the estimates of the denominator degrees of freedom and the scale factor produced by the KR procedure with the restricted method are $n-p$ and 1 respectively.

Proof: By setting $\sigma_i^2 = 0$ for all $i \geq 2$, it can be seen that A_1 and A_2 (as defined in equation 5) reduced to

$$A_1 = w_{11}[\mathrm{tr}(\boldsymbol{\Theta\Phi P}_1\boldsymbol{\Phi})]^2 = \frac{w_{11}\ell^2}{(\sigma_e^2)^2} \qquad \text{and} \qquad A_2 = w_{11}\mathrm{tr}(\boldsymbol{\Theta\Phi P}_1\boldsymbol{\Phi\Theta\Phi P}_1\boldsymbol{\Phi}) = \frac{w_{11}\ell}{(\sigma_e^2)^2}$$

Since $A_1 = \ell A_2$, then the estimates of the denominator degrees of freedom and the scale factor for the KR procedure are $2\ell/A_2$, and 1 respectively (Alnosaier and Birkes, 2019, lemma1).

To compute the estimate of w_{ij}, we used the inverse of the information matrix as in equation (95) on p. 253 of Searl *et all.* (2006).

$$\left(w_{ij}\right)_{i,j=1}^{r} \approx 2\left[\left\{\mathrm{tr}\left(\mathbf{D}_i\mathbf{G}\mathbf{D}_j\mathbf{G}\right)\right\}_{i,j=1}^{r}\right]^{-1}, \tag{9}$$

and

$$w_{11} \approx 2/\mathrm{tr}\left(\mathbf{GG}\right)$$

By straightforward calculation, we can see that:

$$\mathrm{tr}\left(\mathbf{GG}\right) = \frac{1}{(\sigma_e^2)^2}\mathrm{tr}\left[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] = \frac{n-p}{(\sigma_e^2)^2},\ w_{11} \approx \frac{2(\sigma_e^2)^2}{n-p},\ \text{and hence } 2\ell/A_2 = n-p.$$

4. Simulation study

To assess and compare the performance of the KR procedure with the restricted and unrestricted methods to make inference of the fixed effects in mixed linear models, a simulation study was conducted for four different designs:

Design 1: A design with twenty-one observations, nine treatments, seven blocks, and the maximum block size is three, obtained from a partial incomplete block design on p. 329 of Kuehl (2000), and deleting blocks 8 and 9.

Design 2: A design with twenty-one observations, four treatments, nine blocks, and the maximum block size is three, obtained from a balanced incomplete block design on p. 317 of Kuehl (2000), and deleting run 10, and treatment 550.

Design 3: A design with forty observations, obtained from a complete block design with six treatments, seven blocks, and deleting two observations from different blocks and different treatments.

Design 4: A crossover design with four treatments, two periods, and twelve sequences, obtained from section 4.1 of Kenward and Roger (1997).

The model for designs 1, 2 and 3 can be written as

$$y_{ijk} = \mu + \tau_i + b_j + e_{ijk}$$

for $i = 1,...,t,\ j = 1,...,s,\ k = 1,...,n_{ij}$ and all n_{ij} are 0 or 1, where μ is the mean, τ_i is the fixed treatment effects, b_j is the random block effects, and e_{ijk} is the residual errors. The b_j and e_{ijk} are all independent with $b_j \sim \mathrm{N}(0, \sigma_b^2)$ and $e_{ijk} \sim \mathrm{N}(0, \sigma_e^2)$. To have the design matrix of the fixed effects a full column rank to satisfy the assumptions of the model in section 2, the model can be reparametrized by setting $\tau_i^* = \tau_i - \tau_t$ for $i = 1,...,t-1$, and $\mu^* = \mu + \tau_t$. Hence, $\mu + \tau_i = \mu^* + \tau_i^*$ for $i = 1,...,t-1$, and $\mu + \tau_t = \mu^* - \tau_1^* - \cdots \tau_{t-1}^*$. Then the model can be expressed as

$$y_{ijk} = \mu^* + \tau_i^* + b_j + e_{ijk}, \text{ for } i = 1,...,t-1,$$

$$\text{and } y_{tjk} = \mu^* - \tau_1^* - \cdots \tau_{t-1}^* + b_j + e_{tjk} \text{ for } i = t$$

and the null hypothesis of testing that there are no treatment effects in terms of the reparametrized model is $\mathrm{H}_0 : \tau_1^* = \cdots = \tau_{t\text{-}1}^* = 0$.

The model for design 4 can be written as

$$y_{ij} = \mu + s_i + \pi_j + \tau_{k[i,j]} + e_{ij}$$

for $i = 1,...,12,\ j = 1, 2,\ k = 1,...,4$, where μ is the mean, s_i is the random sequence effects, π_j is the fixed period effects, τ_k is the fixed treatment effects, and e_{ij} is the residual errors. The s_i and e_{ij} are all independent with $s_i \sim \mathrm{N}(0, \sigma_s^2)$, and $e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$. Similar to the model for designs 1, 2 and 3, the model is reparametrized so the design matrix of the fixed effects is a full column rank by setting $\mu^* = \mu + \pi_2 + \tau_4$, $\pi_1^* = \pi_1 - \pi_2$, and $\tau_k^* = \tau_k - \tau_4$ for $k = 1, 2, 3$. Then, the terms of the fixed effects in the model can be expressed as:

$$\mu + \pi_1 + \tau_k = \mu^* + \pi_1^* + \tau_k^*,$$

$$\mu + \pi_2 + \tau_k = \mu^* - \pi_1^* + \tau_k^*, \quad \text{for } k = 1, 2, 3,$$

and

$$\mu + \pi_1 + \tau_4 = \mu^* + \pi_1^* - \tau_1^* - \tau_2^* - \tau_3^*,$$

$$\mu + \pi_2 + \tau_4 = \mu^* - \pi_1^* - \tau_1^* - \tau_2^* - \tau_3^*.$$

Then the model is expressed as

$$y_{i1} = \mu^* + s_i + \pi_1^* + \tau_k^* + e_{i1},$$

$$y_{i2} = \mu^* + s_i - \pi_1^* + \tau_k^* + e_{i2}, \text{ for } i = 1,...,12, \text{ and } k = 1,\ 2,\ 3,$$

and

$$y_{i1} = \mu^* + s_i + \pi_1^* - \tau_1^* - \tau_2^* - \tau_3^* + e_{i1},$$

$$y_{i2} = \mu^* + s_i - \pi_1^* - \tau_1^* - \tau_2^* - \tau_3^* + e_{i2} \text{ for } i = 1,...,12, \text{ and } k = 4$$

and the null hypothesis of testing that there are no treatment effects in terms of the reparametrized model is $H_0 : \tau_1^* = \tau_2^* = \tau_3^* = 0$.

For each design, we considered five settings of the ratio ρ, where $\rho = \sigma_b / \sigma_e$ for designs 1, 2, and 3, and $\rho = \sigma_s / \sigma_e$ for design 4, namely $\rho = 0.25,\ 0.5,\ 1,\ 2,\ 4$, and for each setting of ρ, 10,000 data sets have been simulated under the null distribution with no treatment effects. The simulation was done by generating the random terms in each design with each setting of ρ, where, it was assumed that $\mu^* = 0$. Similarly, 10,000 data sets have been simulated under the alternative distribution where there were some treatment effects. In particular, we set $\tau_1^* = 0,$ and $\tau_2^* = \tau_3^* = 0.3$ were added to the generated data sets of designs 2 and 4. For designs 1 and 3, we set $\tau_1^* = \tau_2^* = \tau_3^* = 0,$ and for $i > 3,\ \tau_i^* = 0.3$ were added to the generated data sets..

The REML estimates of the variance components were computed by the iteration algorithm presented in equation (6). The iteration algorithm did not converge for some data sets, and the convergence rates (rounded to one decimal place) were reported in table 1. The convergence rate increased as ρ increased for each design, and it also increased as the design size increased. Almost 100% of all generated data sets for the largest design (i.e., design 3) converged except few data sets for $\rho = 0.25$. For smaller designs (i.e., designs 1 and 2), almost all generated data sets converged for $\rho = 2$ and 4. For the restricted method, the negative variance component estimates set as zero, and the iteration algorithm was re-performed and the REML estimates of the variance of the residual errors were re-computed accordingly.

ρ	Design 1	Design 2	Design 3	Design 4
0.25	89.6	96.3	99.9	99.8
0.5	92.8	97.8	100.0	99.9
1	97.7	99.6	100.0	99.9
2	99.8	100.0	100.0	100.0
4	100.0	100.0	100.0	100.0

Table 1 The percentage of 10,000 generated data sets under the null hypothesis for which the REML iteration algorithm converged

4.1 Instable estimates with the unrestricted method

For some generated data sets, a computational problem was noted with the KR procedure where the procedure was not numerically stable, and the estimates of *m* and λ were outliers. In particular, the computational problem occurred for models with smaller designs (i.e., designs 1 and 2) with smaller values of ρ. In fact, it was noted that the instability occurred with some (not all) of the data sets with negative variance component estimates, and when the estimates allowed to remain negative as suggested by Kenward and Roger (1997) (i.e., with the unrestricted method). Because of the data characteristics of those data sets, it was also noted that the estimate of the quantity A_2 (as defined in section 2) was inflated in such unusual way in which $A_2 > \ell$. This implies that the obtained estimate of the expectation, as in equation 5, $E = 1/(1 - A_2/\ell)$ to be negative (and also the estimate of the variance

V and φ , as in equation 5, were noted to be negative). Typically, the estimates of E and V are expected to be positive since they are the estimates of the expectation and variance of the test statistic F, as in equation 1, which follows approximately the F distribution. The unexpected negative value of the estimate of φ leads to negative value of $(\ell+2)/(\ell\varphi-1)$, and hence unexpected estimate of the denominator degrees of freedom, as in equation 4, $m=4+(\ell+2)/(\ell\varphi-1)$, which is less than 4. In fact, most of the cases the estimate of m found to be negative, and this yields a negative estimate of the scale, as in equation 4, $\lambda=m/[E(m-2)]$. These estimates of the denominator degrees of freedom and scale factor were outliers comparing to the estimates produced by other data sets. In fact, the KR procedure was derived based on matching the first two moments of F distribution which assumes that the denominator degrees of freedom is greater than 4, and the typical estimate of the scale factor to be positive and less than one. When the estimate of the quantity A_2 does not exceed the value of ℓ but so close to it, the quantity $1-A_2/\ell$, which is the denominator of E, becomes close to zero leading to inflated value of the estimate of E. For other data sets with negative variance component estimates, but usual estimate of the quantity A_2, no computational problem was noted.

On the other hand, this numerical instability was not noted with the restricted method for the data sets with negative variance component estimates even when the estimate of the quantity A_2 was inflated. In fact, the estimates of m and λ with the restricted method were not affected by the data characteristics, and were fixed values and equal to $n-p$ and 1 respectively for all generated data sets for all models (see the lemma presented in section 3). To illustrate the computational instability of the KR procedure with the unrestricted method comparing with the restricted method, an example of some data sets is presented below.

Example

Two data sets are considered: a data set (say data 1) generated for the model with design 1, and other data set (say data 2) for the model with design 2. Both data sets were generated under the null distribution, and $\rho=0.25$. The data vectors of 21 observations for each data set are attached in the appendix. The estimates $\hat{\sigma}_b^2$, $\hat{\sigma}_e^2$, $\hat{A}_2$, $\hat{E}$, $\hat{V}$, $\hat{m}$, and $\hat{\lambda}$ were presented in table 2, where $\hat{\sigma}_b^2$ was negative. For the unrestricted method, where $\hat{\sigma}_b^2$ was allowed to remain negative, the estimate of quantity A_2 was inflated for both data sets. It was found that $\hat{A}_2=10.4075>\ell=8$, and $\hat{A}_2=12.1683>\ell=3$ for data 1 and data 2 respectively. Then, $\hat{E}=1/(1-\hat{A}_2/\ell)=-3.1434$, and -0.0439, and $\hat{V}=-0.6844$, and -0.0142 for data 1 and data 2 respectively. Those unusual values of the estimates lead to outlier estimates of $\hat{m}=-4.0132$, and $\hat{\lambda}=-0.2008$ for data 1, and $\hat{m}=-0.1716$, and $\hat{\lambda}=-0.2414$ for data 2. On the other hand, with the restricted method, when $\hat{\sigma}_b^2$ set as zero, the procedure became stable and produced $\hat{m}=12$, and $\hat{\lambda}=1$ for data 1 and $\hat{m}=17$, and $\hat{\lambda}=1$ for data 2. Also, It was noted that when $\hat{\sigma}_b^2$ set as zero, $\hat{\sigma}_e^2$ tended to be underestimated.

The estimating method	Data sets	$\hat{\sigma}_b^2$	$\hat{\sigma}_e^2$	$\hat{A}_2$	$\hat{E}$	$\hat{V}$	$\hat{m}$	$\hat{\lambda}$
The unrestricted method	Data 1	-0.4766	1.4357	10.4075	-3.3230	-0.6844	-4.0132	-0.2008
	Data 2	-0.3692	1.0835	12.1683	-0.3272	-0.0142	-0.1716	-0.2414
The Restricted method	Data 1	0.0000	0.9181	1.3333	1.2000	0.8100	12.0000	1.0000
	Data 2	0.0000	0.6464	0.3529	1.1333	1.1856	17.0000	1.0000

Table 2 The estimates of $\hat{\sigma}_b^2$, $\hat{\sigma}_e^2$, $\hat{A}_2$, $\hat{E}$, $\hat{V}$, $\hat{m}$, and $\hat{\lambda}$ for data 1 and data 2.

4.2 Comparison of the two methods

For the purpose of assessing the performance of the KR procedure with the two methods, and to exclude data sets with outlier estimates, we limited the comparison to *well-behaved* data sets as defined by (Alnosaier and Birkes, 2019) that is the generated data set satisfied: (1) The REML algorithm to estimate the variance components converges, (2) the estimated expectation *E* is positive, and (3) the estimated denominator degrees of freedom $m > 4$. Table 3 presents the percentage of *well-behaved* data sets among the generated data sets (rounded to one decimal place) for the twenty models (four designs for all values of ρ) under the null distribution. It was noted that the percentage of *will-behaved* data sets increased as ρ increased for each design, it also increased as the design size increased. The percentages of *will-behaved* data sets for all designs were almost 100% for $\rho = 2$ and 4. For each of $\rho = 0.25$, 0.5, and 1, the percentages increased as the design size increased. All generated data sets for the model with the largest design (i.e., design 3) were *well-behaved* except only few *ill-behaved* data sets for $\rho = 0.25$.

ρ	Design 1	Design 2	Design 3	Design 4
0.25	72.4	91.4	99.9	99.8
0.50	80.9	95.0	100.0	99.9
1.00	93.8	99.0	100.0	99.9
2.00	99.5	100.0	100.0	100.0
4.00	100.0	100.0	100.0	100.0

Table 3 The percentage of *well-behaved* data sets among 10,000 data sets generated under the null hypothesis

The performance of the unrestricted and restricted methods of the KR procedure was compared based on the level and power of the procedure. The averages of the denominator degrees of freedom estimates, scale factor estimates, and test statistics were also calculated. For each model, the p-value of the test with each method was approximated to be the proportion $\text{Prob}[\text{F}(\ell, m) > \lambda F]$ for each generated data set under the null hypothesis. The observed level of the test is the proportion of these p-values that less than the nominal level 0.05. The adequacy of the method to be evaluated based on the observed percentage of p-values that are less than 0.05, and typically it is desirable to be close to the nominal level of the test 0.05. Similarly, the observed power of the test is the proportion of p-values that less than the nominal level 0.5 under the alternative distribution.

The percentage of *well-behaved* data sets with negative variance component estimates for all models (rounded to one decimal place) are shown in table 4. It was noted that for each design, the percentage of *well-behaved* data sets with negative variance component estimates among the generated data sets decreased as the value of ρ increased. This was expected since for smaller values of ρ, the real variance of the random terms (blocks for designs 1, 2 and 3, and sequences for design 4) were closer to zero, and hence their estimates were closer to be negative. Alnosaier and Birkes (2019) proved in lemma 4 that the residual errors can be simulated with $\mu^* = 0$, and $\sigma_e^2 = 1$, and the other random term simulated with $\mu^* = 0$, and variance equal to ρ. The percentages for the largest design (i.e., design 3) were smaller than for other designs for most values of ρ. In fact, for $\rho = 4$, all *well-behaved* data sets for the four designs had no negative variance component estimates, and typically the two methods are identical. For the purpose of assessing both the methods, and comparing their performance, we considered the designs with the settings of the ratio ρ that had significant numbers of negative variance component estimates, namely $\rho = 0.25$ and 0.5.

ρ	Design 1	Design 2	Design 3	Design 4
0.25	20.1	36.2	37.7	42.2
0.50	15.7	24.0	13.7	28.5
1.00	6.2	5.8	1.2	6.2
2.00	0.6	0.3	0.2	0.2
4.00	0.0	0.0	0.0	0.0

Table 4 The percentage of *well-behaved* data sets among10,000 data sets generated under the null hypothesis and noted to have negative variance component estimates

4.2.1 *The estimates of m,* λ, and the test statistics *with each method*

To compare the performance of the KR procedure with both methods based on the true level, and to get better understanding for the possible differences, we should study the differences of the estimates of the denominator degrees of freedom, the scale factors, and the test statistics since the inner working of the KR procedure depends on these estimates. The average of the estimates of the denominator degrees of freedom *m*, and the average of the estimates of the scale factors λ for the eight models for all *well-behaved* data sets, and for the *well-behaved* data sets with negative variance component estimates were calculated as presented in tables 7 and 8 (in the appendix) respectively. The averages of the estimates of *m* were rounded to one decimal place, and the averages of the estimates of scale factors were rounded to three decimal places.

When all *well-behaved* data sets were considered in the comparison, as presented in table 7 (in the appendix), the average of the estimates of *m* with the restricted method was larger than with the unrestricted method. For each design, the average of the estimates of *m* for $\rho = 0.25$ tended to be larger than for $\rho = 0.5$ with both methods. This was expected since as ρ increases as the information from between blocks (or sequences) utilized to estimate the fixed effects and their variances decreased.

When the comparison was restricted only to the *well-behaved* data sets with negative variance component estimates, it was noted that the average of the estimates of *m* was inflated and larger than for all *well-behaved* data sets with both methods, except model 1 where the estimates almost the same.

Similar to all *well-behaved* data sets, it was noted that the average of the estimates of *m* for the *well-behaved* data sets with negative variance component estimates with the restricted method was larger than for the unrestricted method for all models. In fact, the estimates of *m* for the *well-behaved* data sets with negative variance component estimates with the restricted method were fixed for all data sets of each model, and equal to 12, 17, 34, and 19 for models 1, 2, 3, and 4 respectively, and this coincides with the lemma presented in section 3.

When all *well-behaved* data sets were considered in the computation, the average of the estimates of the scale factors λ were close to 1 for all models, and they were exactly the same or very close to one another for each design as presented in table 8 (in the appendix). The average with the restricted method were equal to 1, and this coincides with the lemma presented in section 3, and this slightly larger than for the unrestricted method for each model by at most 0.011. Similarly, when the comparison was limited to the *well-behaved* data sets with negative variance component estimates, the average with the restricted method was equal to 1 (the lemma in section 3), and except for model 1, this was slightly larger than for the unrestricted method for all models by at most 0.005. For the unrestricted method, the average of the estimates of λ for model 1 (small design size with 9 treatments fixed effects) was the smallest average, 0.959, and it was equal to 1 for the model with largest design (i.e., design 3).

The average of the test statistics for all *well-behaved* data sets, and also for the *well-behaved* data with negative variance component estimates was presented in table 9 (in the appendix). The averages of the estimates were rounded to three decimal places. It was noted that the test statistic was underestimated for the *well-behaved* data sets with negative variance component estimates. However, with the restricted method, and by setting the negative variance components as zero, $\hat{\sigma}_e^2$ tended to be underestimated (as illustrated in the example of section 4.1), and also, it is worth mentioning that the test statistic is scaled by λ which is 1 with the restricted method. This inflated the test statistics for the restricted method comparing to the unrestricted method. On other words, the test statistic for the *well-behaved* data sets with negative variance component estimates found to be underestimated, but with the restricted method tend to be less underestimated than with the unrestricted method.

4.2.2 Comparison of the test level for each method

The observed percentage of p-values that were less than 0.05 under the null distribution for all *well-behaved* simulated data sets, and for the *well-behaved* data sets with negative variance component estimates for the 8 models (4 designs with two values of ρ) with both methods are presented in table 5. Typically, because of the simulation error, the observed percentage of p-values was not expected to be exactly 0.05, and the simulation error is $22\%/\sqrt{n}$ where *n* is the number of data sets considered in the computation.

When all *well-behaved* simulated data sets considered in the comparison, it was noted that 13 out of 16 of the observed percentages of p-values were liberal (i.e., significantly larger than 0.05). For the model with the largest design (i.e., design 3), there was no significant difference between the observed percentage and 0.05 with both methods for the case $\rho = 0.5$. Assuming that 0.2% as unimportant difference between two observed percentages, the KR procedure "appeared" to perform similarly for both methods based on the test observed level, for 3 models out of 8. For the other five models, the

procedure with the unrestricted method "appeared" to perform more adequately. Moreover, the observed percentage was more liberal for all *well-behaved* data sets than for *well-behaved* data sets with negative variance component estimates.

However, when restricting the comparison of the two methods to the only the *well-behaved* data sets with negative variance component estimates, the situation reversed in some ways. It was noted that the observed percentage with the unrestricted method was excessively conservative for all models except the crossover models. On the other hand, except the models with designs 2 and 3 for $\rho = 0.5,$ the performance of the restricted methods was adequate where there was no significant difference between the observed percentage and 0.05. For models with designs 2 and 3, for $\rho = 0.5,$ the observed percentage was conservative with the restricted method; however, it was not excessively conservative as for the unrestricted method. In fact, this explains why the observed level with the unrestricted method "appeared" to be less inflated than for the restricted method when all *well-behaved* data sets above was considered in the comparison.

The excessively conservative observed level for the *well-behaved* data sets with negative variance component estimates for the unrestricted method are due to the underestimated test statistics as seen in section 4.2.2, and seemed not affected by the larger estimated denominator degrees of freedom. For the restricted method, as seen in section 4.2.2, the test statistic found to be less underestimated, and the estimated degrees of freedom larger. This yielded less excessively conservative observed level. The KR procedure with the restricted method found to be more adequate for the models with block designs based on the observed level. For the model with the crossover design, both methods found adequate and performed similarly.

The estimating method	ρ	All *well-behaved* data sets				*well-behaved* data sets with negative estimates of variance components			
		Design 1	Design 2	Design 3	Design 4	Design 1	Design 2	Design 3	Design 4
The unrestricted method	0.25	6.54	5.29	5.46	6.27	0.78	2.60	4.24	5.10
	0.50	6.15	5.76	5.10	5.75	0.72	2.76	2.62	4.95
The Restricted method	0.25	7.27	6.11	5.84	6.13	5.30	4.84	5.25	4.77
	0.50	6.59	6.16	5.22	5.79	4.64	4.52	3.50	5.09

Table 5 The observed percentage of p-values that were less than 0.05 under the null hypothesis for all *well-behaved* simulated data sets and *well-behaved* data sets with negative estimates of variance components

4.2.3 Comparison of the test power with each method

To compare the performance of the two methods based on the test power, the percentage of p-values that were less than 0.05 under the alternative distribution for all *well-behaved* data sets, and the *well-behaved* data sets with negative variance component estimates for both methods are presented in table 6. When all *well-behaved* data sets were considered in the comparison, it was noted that the test was more powerful with the restricted method. This was expected because of the so liberal observed test level with the restricted method as noted in section 4.2.2. The test for $\rho = 0.5$ was significantly more powerful than for $\rho = 0.25$ for both methods, and the model with largest design (i.e., design 3) had the most powerful test with the restricted method, 94.88%. Also, except the models with the smaller

designs (i.e., design 1 and 2) with $\rho = 0.5$, the test tended to be more powerful when all *well-behaved* data sets considered in the computation than for the *well-behaved* data sets with negative variance component estimates. This was also expected based on our findings in section 4.2.2 that the observed level for all *well-behaved* data sets was more liberal than for *well-behaved* data sets with negative variance component estimates for each model with both methods.

Similarly, when restricting the comparison to the *well-behaved* data sets with negative variance component estimates, the test was more powerful with the restricted method for all models. This was expected because the test was excessively conservative with the unrestricted method as seen in section 4.2.2. The test, as expected, for models with large designs found to be more powerful, and with $\rho = 0.5$ was more powerful than with $\rho = 0.25.$ for all models, the KR procedure found to be more powerful with the restricted method.

The estimating method	ρ	All *well-behaved* data sets				*well-behaved* data sets with negative estimates of variance components			
		Design 1	Design 2	Design 3	Design 4	Design 1	Design2	Design 3	Design 4
The unrestricted method	0.25	21.82	26.11	35.66	34.97	5.94	18.64	28.99	28.29
	0.50	70.7	78.27	94.60	88.99	48.93	69.40	91.42	84.13
The Restricted method	0.25	27.62	30.35	37.57	38.65	27.47	29.30	34.02	37.00
	0.50	78.25	81.85	94.88	91.09	84.30	83.57	93.44	91.57

Table 6 The observed percentage of p-values that were less than 0.05 under the alternative hypotheses for all *well-behaved* simulated data sets and *well-behaved* data sets with negative estimates of variance components

5. Discussions and conclusions

Like most of variance component estimation methods, the (REML) method sometimes produces negative estimates, although variances should be positive by definition. This is true because the method equations do not guarantee the estimates to be non-negative. In the simulation study, it was found that the negative variance component estimates occurred more often with smaller values of the ratios of the variance components ρ and small size designs. The negative variance component estimates in normal mixed linear models found to have an impact on the performance of the KR test for the inference of fixed effects since they affected the estimates of the denominator degrees of freedom, the scale factors, and the test statistics, and eventually affected the test level and power. The KR procedure with the restricted method found to overcome the procedure with the unrestricted method in two ways. First, the procedure with the unrestricted method found to be instable for some data sets with negative variance component estimates, especially for models with small size designs. Second, when the procedure was stable, the procedure with the restricted method found to perform better based on the test level and power.

When the KR procedure with the unrestricted method was instable, the procedure produced outlier estimates for both the denominator degrees of freedom and scale factor, and the procedure performed poorly. When the procedure was stable with the unrestricted method, the estimated denominator degrees of freedom tended to be larger with negative variance component estimates, and the scale

factor tended to be slightly smaller. On the other hand, the test statistic was underestimated. Because of the underestimated test statistics, the observed level of the test was excessively conservative for the data sets with negative variance component estimates, and it appeared not affected significantly by the enlarged estimated degrees of freedom. Typically, the observed power became less powerful with negative variance component estimates. For the crossover design in the simulation study, the test was adequate based on the test level. However, the test was so excessively conservative for the block designs.

By setting the negative variance component estimates as zero with the restricted method, the KR test found to be stable for data sets that the procedure was instable with the unrestricted method. By setting the negative variance component estimates as zero, the simulation study showed that the other variance component estimates (i.e., the variance of the residual error) was deflated as expected. This altered the test statistic values, the estimates of the denominator degrees of freedom, and the scale factors, and eventually altered the observed level and power of the test. The denominator degrees of freedom and scale estimates found to be fixed and not rely on the data character. In fact, the test statistic became larger than with the unrestricted method (i.e., less underestimated). The larger estimate of denominator degrees of freedom, scale of 1, and larger test statistic made the observed level less excessively conservative than with the unrestricted method, and eventually the test tended to be more powerful.

For the crossover design, the test with the restricted method was as adequate as with the unrestricted method based on the test observed level, however, the test was more powerful with the restricted method. For the block designs, the observed level with the restricted method was less excessively conservative and the test was more powerful.

The adequacy of the restricted method for the KR test found here might be limited to the similar designs presented in this paper, and more studied and investigation for the KR test with both methods for other designs is suggested.

Appendix

Data sets for the example in section 4

The data rounded to four decimal places, and they are in blocks order.

Data 1: [0.8773, -0.2444, 1.2805, 0.3426, -0.2859, -0.2334, 0.8470, -0.1763, 0.1673, -0.3426, 0.5562, 0.7307, 0.9014, 0.6246, 0.3213, -2.043, 1.0356, 0.9406, 1.3180, -0.8350, -0.0956]

Data 2: [-0.750, 0.7134, 0.3750, 0.2506, -0.1666, -0.1719, -0.4631, -1.2781, -0.4384, -1.0419, -0.4703, 1.5861, -0.524, -0.8910, 0.3684, 0.3383, -0.3487, -0.0774, 1.6013, 0.6131, 0.9513]

The estimating method	ρ	All *well-behaved* data sets				*well-behaved* data sets with negative estimates of variance components			
		Design1	Design2	Design3	Design 4	Design1	Design2	Design3	Design4
The unrestricted method	0.25	7.8	13.4	28.5	14.4	7.5	15.0	28.8	16.3
	0.50	7.7	12.8	28.3	13.6	7.5	15.1	28.7	16.3
The Restricted method	0.25	9.6	14.3	30.5	15.5	12.0	17.0	34.0	19.0
	0.50	9.0	13.4	29.0	14.4	12.0	17.0	34.0	19.0

Table 7 The averages of the estimated denominator degrees of freedom for all *well-behaved* simulated data sets and *well-behaved* data sets with negative estimates of variance components.

The estimating method	ρ	All *well-behaved* data sets				*well-behaved* data sets with negative estimates of variance components			
		Design1	Design2	Design3	Design4	Design1	Design2	Design3	Design4
The unrestricted method	0.25	0.989	0.998	1.000	0.992	0.959	0.995	1.000	0.988
	0.50	0.992	0.999	1.000	0.994	0.959	0.996	1.000	0.988
The Restricted method	0.25	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000

Table 8 The averages of the estimated scales for all *well-behaved* simulated data sets and *well-behaved* data sets with negative estimates of variance components.

The estimating method	ρ	All *well-behaved* data sets				*well-behaved* data sets with negative estimates of variance components			
		Design1	Design2	Design3	Design4	Design1	Design2	Design3	Design4
The unrestricted method	0.25	1.4 77	1.247	1.079	1.277	0.875	0.944	0.976	1.115
	0.50	1.485	1.244	1.082	1.284	0.900	0.956	0.961	1.126
The Restricted method	0.25	1.496	1.301	1.103	1.276	1.210	1.108	1.038	1.111
	0.50	1.498	1.272	1.088	1.284	1.207	1.090	1.013	1.125

Table 9 The averages of the test statistics for all *well-behaved* simulated data sets and *well-behaved* data sets with negative estimates of variance components.

References

1- Alnosaier W, Birkes D (2019) Inner workings of the Kenward-Roger test. Metrika 82:195-223

2- Corbiel RR, Searl SR (1976) Restricted maximum likelihood (REML) estimation of variance components in the mixed model. Technometricso 18(1): 31-38

3- Gao B, Li S, Li W, Li S, Wang X (2008) Statistical analysis of negative variance components in the estimation of variance components. VI Hotine-Marussi Symposium on Theoretical and Computational Geodesy. International Association of Geodesy Symposia, vol 132. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74584-6_48

4- Hartley HO, Rao JN (1976) Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54: 93-108

5- Harville DA (2008) Matrix algebra from a statistician's perspective. Springer, New York

6- Harville DA, Jeske DR (1992) Mean squared error of estimation or prediction under a general linear model. J Am Stat Assoc 87:724-731

7- Kackar RN, Harville DA (1984) Approximation for standard errors of estimators of fixed and random effects in mixed linear models. J Am Stat Assoc 79:853-862

8- Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983-997

9- Kuehl RO (2000) Design of experiments: statistical principles of research design and analysis, 2nd edn. Duxbury Press, Pacific Grove.

10- Littell R, Milliken G, Stroup W, Wolfinger R, Schbenberger O (2006) SAS for mixed models. SAS Institute Inc., Cary

11- SAS Institute Inc (2021) SAS/STAT15.2 User's guide. SAS Institute Inc., Cary

12- Searl SR, Casella G, McCulloch CE (2006) Variance components. Wiley, New York

13- Searl SR, Gruber MH (2016) Linear models. John Wiley & Sons, New York

14- Sjoberg L (1995) The best quadratic minimum bias non-negative definite estimator for an additive two variance component model. Manuscr, Geodaet, 20(2):139-144

15- Stroup WW (1999) Mixed model procedures to assess power, precision, and sample size in the design of experiments. American Statistical Association 1999 Proceedings of the Biopharmaceutical Section: 15-24

16- Stroup WW, Littell RC (2002) Impact of variance component estimates on fixed effect inference in unbalanced linear mixed models. Conference on applied Statistics in Agriculture

17- Verdooren LR (1982) How large is the probability for the estimate of a variance component to be negative? Biom J. 24(4):339-360

18- Wu X (1992) Helmert-WF (weight factor) variance component estimation. J. Wuhan Tech. Univ. Surv. Map 17(4):1-10

19- Zhang F, Feng C, Huang C, Li Y-Q (2000) Application of the improved Helmert method for variance component estimation in precise orbit determination. Act Geod. Cartogr, Sinica 29(3): 217-223